

國立臺灣師範大學特殊教育學系、特殊教育中心
特殊教育研究學刊，民 91，23 期，121-140 頁

噪音背景辨識語音測驗編製之研究

蔡志浩

陳小娟

高雄市立民族國小

國立高雄師範大學

本研究旨在發展一份噪音背景辨識語音測驗。以年齡介於 20 至 30 歲之聽力正常年輕人為研究對象，樣本為三個部分：預測度分析 12 人、清晰度分析 31 人、複本測驗的發展 30 人。

測驗編製過程包括預測度分析、清晰度分析、語音與字詞熟悉度分析、信度分析與複本測驗的發展、效度分析等。研究結果摘述如下：

1. 預測度分析：測驗中所有高預測度句子之預測度分數介於 1 至 11 分間，具有一定的預測度。

2. 清晰度分析：採 0 dB 與 -2 dB 兩種信噪比，高預測度句子之清晰度為 19 或 20 分、低預測度句子的清晰度介於 1 到 16 分間、高低預測度句子之得分差距則介於 3 至 19 分之間。

3. 語音成分分析：150 個關鍵字音節結構均不重複，無論是 300 個句子或各表單 50 個句子，其整體語音成分與國語語音成分都很接近。

4. 熟悉度分析：150 個關鍵字之口語出現頻次介於 1 至 1300 次之間，所有的關鍵字均為國小學童認識的字並屬於常用字範圍。

5. 測驗的信度分析與複本測驗的發展：本測驗共計發展了六張不同的表單，這六張表單在高預測度句子、低預測度句子及高低預測度句子得分差距之分數上，彼此相關程度很高，顯示這六張表單具有複本信度，亦說明六張表單彼此互為複本。

6. 效度分析：經預測度、清晰度、熟悉度與語音成分等幾項分析，並經專家審查，顯示本研究具有一定的內容效度。六個表單進行因素分析後，分別得到高預測度句子與低預測度句子兩項因素，顯示本測驗具有構念效度。而以驗證性因素分析檢驗六個表單互為複本的適合度，結果亦顯示六個表單彼此互為複本。

本研究並根據上述研究結果，提出未來研究方向的建議。

關鍵字：噪音、語音辨識測驗、預測度、清晰度、熟悉度

緒論

一、研究動機

說話理解是一個複雜的歷程，包括對語音的覺知與辨識，並成功地整合字、詞、句子成為有意義的訊息（Pichora-Fuller, Schneider & Daneman, 1995）。日常生活中存在著許多種不同的噪音，當說話訊息受到噪音的影響，訊息特別容易消逝，這時聽者就得運用額外的策略去理解訊息，而運用文意脈絡訊息就是其中一項策略，Elliott 將這種在噪音中運用文意脈絡訊息去理解說話的能力稱為聽覺閉鎖能力（auditory closure）（Elliott, 1995）。

聽力損失者由於聽力損失，接收到的是扭曲且片斷的訊息，一旦加上噪音的干擾，往往造成聽力損失者抱怨在噪音情境下對說話訊息的理解感到困難，雖然有這樣抱怨事實的存在，然而純音聽力檢查及一般的語音聽力檢查由於是在安靜情境下施測，因此所測出之結果難以預測聽力損失者在噪音中對說話訊息的理解（Dubno, Dirk & Morgan, 1984; Elliott, 1995; Plomp, 1986），以致於聽力損失者即使配戴助聽器，仍感到不滿意，持續使用的意願不高（Agnew, 1997）。如果在溝通能力的評量中不考慮這個部分，只注意到個體在安靜環境中接收訊息的情形，往往會錯估了其真實表現。

Kalikow, Steven 和 Elliott 於 1977 年接受美國國家神經疾病與中風機構（National Institute of Neurological Disease and Stroke，簡稱 NINDS）的委託，發展了噪音背景辨識語音的測驗（Speech Perception In Noise，本文簡稱 SPIN），這項測驗是以一種 12 個人同時說話的雜踏式噪音（babble noise）為背景，測驗聽者在高預測度句子（high-predictability sentence，本文簡稱為 HP 句子）及低預測度句子（low-predictability sentence，本文簡稱為 LP 句子）中運用文意脈絡（semantic-contextual）訊息及語音音響（acous-

tic-phonetic）訊息的能力。

SPIN 測驗發展後，在年輕聽力正常者、聽力損失者、老年人身上得到很好的信度及效度，並廣泛地被應用在臨床上（Bilger, Nuetzel, Rabinowitz & Rzeczkowski, 1984; Cokely & Humes, 1992; Elliott, Lyons & Busse, 1983; Elliott, 1995; Huchinson, 1989; Kalikow et al., 1977; Jerger, Oliver & Pirozzolo, 1989; Morgan, Kamm & Velde, 1981; Wild & Humes, 1990）。SPIN 測驗的結果不但能提供聽力損失者在日常生活溝通能力的訊息；在助聽器的選配時，亦可以用來比較不同款式助聽器的效益以及了解配戴助聽器後在噪音中聽取語音訊息的能力（Agnew, 1997; Elliott, 1995; Hank & Johnson, 1998; Kalikow et al., 1977）。此外，SPIN 測驗亦可應用在外語學習（Mayo, Floretine & Buus, 1997）、在噪音中有學習困難的學生（Jerger, 1996）、非聽覺性輔助器材（Kalikow et al., 1977）、與聽力保護用具（Wilde & Humes, 1990）等研究上。

日常生活中說話的題材很廣泛且在不同背景噪音的情境中出現，所以不可能以單一測驗去推論多種日常生活事件中語音的聽取能力，最好的方法是採用不同的語音測驗，並且設計一種能評量不同障礙程度的測驗（Kalikow et al., 1977）。目前國內雖然發展了一些語音辨識測驗，但都是在安靜中測試，欠缺在噪音下的語音測驗，研究者期望能藉由發展中文版的 SPIN 測驗，探討聽力正常成年人在噪音背景中運用文意脈絡訊息及語音音響訊息的能力，做為提供未來應用在聽力損失者語音聽知覺相關研究之參考。

二、研究目的與待答問題

Kalikow 等（1977）歸納文獻指出一項好的語音測驗應同時考量下列項目，並依據這些考量編製了 SPIN 測驗。（一）語音及超語段因素（phonetic and prosodic factors）：以句子為材料，包括語音及超語段兩種訊息，除了能反應

真實說話的情境，也有助於聽者理解訊息。(二) 句子脈絡效應 (effect of sentences context)：句子可以用來評量前後文意脈絡，若只是採用單字便沒有前後脈絡可言。(三) 字詞的熟悉度 (word familiarity)：字詞的熟悉度會影響訊息接受的清晰度。(四) 噪音干擾 (noise interference)：日常生活是在噪音環境中接受訊息而不是在安靜的環境之中。(五) 聽者本身相關因素 (listener-related factors)：除了考量聽力損失者與聽力正常者在噪音干擾下接受訊息的不同，也考量不同聽力損失者對於文意脈絡及詞彙數量之運用能力也可能不同。(六) 其它語音辨識測驗的問題 (problems with existing text of speech perception)：對受試者而言，一次辨識整個句子，牽涉的認知歷程較複雜，且施測時間較長；對施測者而言，整個句子的計分方式也較複雜，如果縮小辨識單位，可能比辨識整句為佳。

本研究期望能依據同樣的原則，編製符合國語語言特質的噪音背景中語音辨識測驗，並瞭解聽力正常成人之本測驗的表現。

本研究依據研究目的提出以下待答問題：

(一) 句子編製

1. 高預測度句子在紙筆預測度測驗中的預測度如何？

2. 高預測度句子在雜踏式噪音背景中的清晰度如何？

3. 低預測度句子在雜踏式噪音背景中的清晰度如何？

4. 各表單關鍵字之熟悉度為何？

5. 各表單句子之語音分布情況為何？

(二) 測驗的複本信度

1. 不同表單中高預測度句子的複本信度如何？

2. 不同表單中低預測度句子的複本信度如何？

3. 不同表單中高預測度句子與低預測度句

子間差距分數的複本信度如何？

(三) 測驗的效度

1. 測驗的內容效度如何？

2. 測驗的構念效度如何？

3. 在驗證性因素分析中，不同表單之高預測度句子得分彼此是否互為複本？

4. 在驗證性因素分析中，不同表單之低預測度句子得分彼此是否互為複本？

5. 在驗證性因素分析中，不同表單之高預測度句子與低預測度句子得分差距彼此是否互為複本？

研究方法

一、句子的編製

(一) 確立句型編製規則 (Formulation of an initial corpus of sentence)

依本研究目的，研究者編製 HP、LP 兩種句子，其編製的規則如下：(1) 高預測度句子：每句包括 8 至 10 個字，句子中有 2 至 3 個預測句末關鍵字的線索，例如：水裡有隻游來游去的魚。(2) 低預測度句子：每句包括 8 至 10 個字，句子中沒有預測句末關鍵字的線索，例如：我不知道還有沒有水。

(二) 關鍵字的選取

1. 語音音節不重複的考量

本研究依國語日報字典 (民 85) 所列，統計出國音 21 個聲母和 16 個韻母所有可能組合的音節結構共 414 個，而這些音節結構再加上聲調變化共有 1313 個。

2. 熟悉度的考量

在 1313 個音節中，每個音節選取至少一個以上相對應的同音國字，並請三位國小高年級小朋友判斷 (1) 是否學過這個字、(2) 這個字是否是他常用的字，合乎這兩項條件的國字則列入編製句子可能採用的關鍵字。經熟悉度考量後所留下之關鍵字共計 1002 個字。

3. 文意脈絡的考量

依據熟悉度選取關鍵字後，再依 HP 句子的造句規則，編製 HP 句子。而若有無法符合造句規則、或想不出來之情形，則捨棄這個國字。經由文意脈絡的考量共計留下 2102 句。

(三) 句子的難度及合理性的考量

每個句子在編製後，請 3 位國小高年級小朋友及 3 位中小學老師依據：1. 理解這個句子的句意是否有困難、2. 在日常生活中使用這個句子是否合理等兩項標準判斷，若未符合以上兩項條件者則刪除此句。最後並請國內聽語專家，高師大特教系陳小娟教授進行審查。經由句子難度與合理性的考量後，共計留下 2053 句。

(四) 預測度分析

經上述過程後留下高預測度句子 2053 題，依隨機排列的方式呈現，本研究採 Kalikow 等 (1997) 之相同預測度分析方式，請 12 位年齡在 20 到 30 歲間之聽力正常年輕人進行紙筆預測度測驗。

統計十二位受試者每個句子的得分，以之為預測度，每句答對得一分，因此每句之得分最低零分，最高為十二分。預測度測驗的目的是刪除預測度太高的句子（即 12 分的題目）及預測度太低的句子（0 分的題目）。此外，為顧及音節不重複原則，每個音節結構至少保留一句。經由這兩項標準所留下來的題目共計 370 題。另外再依據 370 題 HP 句子的關鍵字編製 370 題相對應之 LP 句子，這些題目同樣請三位國小學生及三位中小學教師進行難度與合理性的考量，並進行題目修改，最後再請陳小娟教授審查。在 370 個 LP 句子中，有 18 句用的是同一個關鍵字，因此在審查時，由兩句中擇取編製較佳的一句，共刪除了其中 18 個 LP 句子，最後共計留下 352 個 LP 句子。

二、雜踏式噪音的製作

雜踏式噪音 (babble noise) 是錄製六位發

音清晰之成年人（三男、三女）朗讀兒童故事書中之不同段落，再經編修而成。錄音地點位於高師大特教系一樓聽力檢查室 (IAC 140)；錄音時，麥克風 (JVC MV-88 型) 距離口部 10 公分並與口部成零度平行角，以避免發音氣流直接撞擊麥克風所造成不必要的噪音和扭曲音，並用視覺監控 CSL 音量指示器，以期維持相同的音量。麥克風連接在 KAY CSL (Computerized Speech Lab, 以下簡稱 CSL) 4300B 的語音分析編輯系統硬體的連接埠上，而 KAY CSL 4300B 之語音編輯軟體則安裝在 486DX2-66 的個人電腦內。本研究所設定的取樣數率 (sampling rate) 為 20kHz，數位低通濾波器通過頻率為 8kHz，使用解析度 (resolution) 為 16 位元的數位/類比 (AD/DA) 信號轉換卡。將輸入語音存在檔案中，之後以該儀器語音編輯軟體處理。

語音處理之過程，首先是刪除句子與句子間超過 0.04 秒之過長的空白時段，之後截取每人一分鐘的朗讀段落。再將每位說話者錄音長度分為前後兩半段（每段各 30 秒），然後以混音方式處理。

混音過程採以下程序實施：

1. 前半段混音程序：

(1) 先把六位說話者前半段說話的音量，兩人一組共編為三組，將每一組兩個人各以一分鐘長度之語音以相同音量 (60 dB) 進行第一次混音，如此混合成三組第一次混音。

(2) 選取第一次混音中的兩組混音（第一組和第二組），以相同音量 (60 dB) 再進行第二次混音，得到第四組混音。

(3) 將第一次混音但未進行第二次混音的那一組混音（第三組）以及剛混音完成的第四組都調整成相同音量大小 (60 dB) 後，進行第三次混音，如此就得到一組六個人同時說話的混音（第五組）。

2. 後半段混音程序：與前半段混音程序相

同，經過第六、七、八、九組的混音，同樣也可得到一組六個人同時說話的混音（第十組）。

3. 最後將前半段混音（第五組）與後半段混音（第十組）調整成相同音量（60 dB），進行最後一次混音（第十一組）。

分析計算整體聲音的平均音量後，接著測量噪音整體的變動量。在取樣時間間隔為 0.1 秒下，百分之九十時間內整體音量變動都在 ± 4 dB 範圍內，測量目的是在確認整體雜踏式噪音在每段時間的音量是否穩定，若符合此項標準，才算是完成雜踏音的製作程序。

三、清晰度分析

(一) 測驗句子的錄製

以紙筆預測度分析過程所得到的 370 個 HP 句子與 352 個 LP 句子做為清晰度測驗材料，這 722 個句子請一位國語標準且說話清晰的成年男子做為主講者，錄音地點與設備規格同前所述。

語音處理的步驟如下，在 CSL 其中一個軌道上播放題號，例如「第一題」，接著是 6 秒的雜踏式噪音，當雜踏式噪音播放了 2 秒後，另一軌則開始播放句子，句子結束 2 秒後，停止播放噪音。噪音結束至下一題題號出現有 4 秒間隔，加上句子結束後的 2 秒噪音，因此受試者有 6 秒作反應。

(二) 在噪音中句子清晰度的分析

以年齡介於 20 至 30 歲之聽力正常年輕人做為清晰度測驗之樣本。受試者篩選條件如下：1. 聽力正常：250Hz、500Hz、1000Hz、2000Hz、3000Hz、4000Hz、6000Hz、8000Hz 等頻率之純音聽閾均為 20 分貝（含）以下。2. 在 3000Hz、4000Hz 與 6000Hz 沒有凹谷（所謂凹谷，是指上述三個頻率聽閾比左右相鄰頻率之聽閾大，而且相差至少 10 分貝）。3. 中耳功能測驗正常，亦即鼓室圖為 A 型，而且 500Hz 與 1000Hz 有對側聽小骨肌反射現象。合乎上

列條件的共計 31 人，包括 9 位成年男子、22 位成年女子，平均年齡為 24 歲。

清晰度分析目的除了要選出清晰度較同質的句子，也要選出 HP、LP 句子得分差異較大的句子。在正式研究之前，本研究依據 Hucherson, Dirk 和 Morgan (1979) 不同信噪比（-4 dB, -2 dB, 0 dB, 2 dB, 4 dB, 6 dB, 8 dB）之 HP 與 LP 句子得分的研究發現，選取 4 dB、2 dB、0 dB、-2 dB 及 -4dB 等 5 個信噪比，先請兩名聽力正常成年人進行一項預試。這兩名受試者在 4 dB 及 2 dB 兩個信噪比之 HP 句子得分接近 100%，而在 -4 dB 信噪比之下，LP 句子得分接近 30%。於是排除 HP 句子得分過高（接近 100%）及 LP 句子得分過低（接近 30%）之信噪比，而以 0 dB 及 -2 dB 做為正式清晰度測驗之信噪比。

正式清晰度測驗是將雜踏式噪音固定為 70 分貝，測試 31 位聽力正常年輕人在每句的清晰度分數。施測地點位於高師大特教系聽力檢查室，將測驗題目存於語音編輯軟體檔案中，經 CSL 連接 Beltone 2000 聽力檢查計，以 TDH-39 耳機為輸出體，在優耳呈現題目。施測前由 Beltone 廠商校正聽力檢查計。

清晰度分析的 722 個句子依據「音節不重複」、「高低預測度句子在不同的兩張表單出現」、「高、低預測度句子不可以連續出現三次」等原則，以半隨機方式分別派入不同的表單，每張表單有 50 句，計 16 張表單（其中有兩張表單只有 11 題）。

由於清晰度測驗的句子共計 722 題，施測音量為 70dBHL，而信噪比為 0 dB 與 -2 dB，若同一位受試者要聽完所有測驗題目，兩種信噪比情境約需 8 小時。如此冗長時間參與測驗，不但會增加受試者的疲勞度，也會降低其再度參與之意願。因此，本研究採用對抗平衡設計，以參與次數做為受測單位。每次的參與（一個受測單位）相等於一位受試者在兩種信

噪比情境，各聽取八分之一的題目，因為有兩種信噪比，所以每次參與可聽取四分之一的題目；結合四個人次（即四個受測單位）即是一個完整的測驗。受試者中有 9 人參與兩次，其餘 22 人則各參與一次，參與次數總計 40 次，亦即完成 10 次完整的清晰度測驗。換言之，每個題目有 20 個人次作答（其中 0 dB 與 -2 dB 之情境各 10 人次作答）。每次測驗時間約為兩小時，測驗結束後致贈兩百元禮券做為酬勞。

清晰度分析測驗的計分方式，是計算每個句子在 10 次 0 dB 及 10 次 -2 dB 兩種信噪比之下的總分。每題一分，最高分為 20 分，最低分為 0 分。

清晰度分析主要目的是提高 HP 和 LP 的句子同質性，並擴大 HP 和 LP 句子得分之差異。因此本研究依據 Kalikow 等（1977）清晰度分析之刪題標準，刪除 HP 得分太低的句子及 LP 得分太高的句子。且同一個關鍵字至少仍保留 HP 與 LP 句子各一句。同時，在考量音節不重複原則，本研究在每個音節至少保留一組相對應之 HP 與 LP 句子。依據受試者在清晰度分析的得分與上述幾個刪題原則，本研究選取 HP 句子清晰度得分在 19 分（含）以上而 LP 句子清晰度得分在 16 分（含）以下的題目。經此過程，HP 句子留下 184 題，而 LP 句子亦留下 184 題，合計有 368 個句子。

四、語音與熟悉度分析

（一）音素分析

本研究進一步分析經清晰度分析後留下的 368 個句子其關鍵字的音節結構，依每一種音節結構每張表單只選取一個字，而且每張表單有 50 個關鍵字的原則，共組成六張表單，合計有 300 個題目，這 300 題含有的 150 個關鍵字其音節結構皆不重複。

（二）測驗句子之語音成分分析

測驗句子的語音成分分析共分為兩部分：第一部分是分析整體 300 個句子之語音成分。

第二部分則是分析各表單 50 個句子之語音成分。

（三）關鍵字熟悉度分析

本研究依據教育部（民 88）八十七年口語調查報告書之口語統計字頻表，計算每個關鍵字的出現頻率。

五、信度分析與複本測驗的發展

（一）複本測驗的選擇及錄音

經清晰度測驗留下的句子分別派入六個表單。每張表單具有以下特質：1. 每個表單有 50 句，其中 HP 與 LP 句子各 25 句；2. 每一個奇數與偶數的表單含有相同的關鍵字，但是這些字卻是放在相反預測性的句子（HP、LP 句子）裡；3. HP 和 LP 句子隨機排列，每個表單內 50 個關鍵字均不重複；4. 每一個表單和其它的表單相似（其相似的部份包括：語音成份相似、HP 與 LP 句子的平均清晰度相似、HP 句子的預測性相似、字詞熟悉度相似）。

將每張表單進一步做以下的檢查：確定每一個表單其清晰度相近、平均語音成分相近、預測度相近。檢查方式如下：若這 50 句組成的表單，其清晰度與預測度分數之平均與標準差在不同表單差異很大，就和其它表單的句子更換，目的是要取得平衡，以維持每張表單有相近的清晰度與預測度。所以在調換句子的時候，同時考量了清晰度、預測度及語音分布的平衡。

每張表單再依以下兩個原則重新重組：1. 以半隨機的方式排列：不可以三種同預測性的句子連續排列在一起（如 HP 句子不可以連續出現三次以上）；2. 排在一起的句子，在意思上不可以是相關的，如果相鄰的句子彼此有語意相關，無論是 HP 或 LP 句子，都要再重新排列。

最後，以單因子變異數考驗六個表單中 HP 句子清晰度、LP 句子清晰度、預測度等三項分數是否有顯著差異。

這六張表單再由同一位成年男子錄音。錄音地點、器材、方式與語音處理同前所述，即語音錄在一個軌道上，雜踏式噪音錄在另一個軌道上，並在每個表單一開始時錄製一段 1kHz 的校準音。

(二)複本型式的考驗

受試者為 30 位聽力正常年輕人，聽力檢查項目與標準與清晰度分析相同，呈現刺激音於優耳，每張表單的信噪比為 -2 dB，信號及噪音的音量分別是 68 dBHL 與 70 dB HL。選擇信噪比為 -2 dB 的原因是參考國外研究 (Elliott, 1979; Kalikow et al., 1977; Morgan, Kamm & Velde, 1981) 與本研究清晰度分析之結果後，選取 HP-LP 的差距分數最大的信噪比 -2 dB。施測地點與所用儀器同前所述。

1. 指導語

施測時將指導語錄於測驗題目之前，指導語如下：「在您的前面有一張作答的表單，表單內共有 50 題。等一下您會在一耳先聽到題號，兩秒之後，會有一個伴隨著吵雜噪音的句子出現，請仔細聽並依據題號用筆寫下每一個句子的最後一個字，若忘了這個字怎麼寫，可用注音符號作答，做為評分者判斷的依據，現在先請您聽練習題，之後再聽正式句子。若有問題或不清楚的地方，可請施測人員為您再解釋一次。」。

2. 施測程序

受試者坐在聽力檢查室內，先聽兩句練習題，在每句開始說話前兩秒先聽題號「第一題、……」，題目與題目間相隔 4 秒，加上句子結束後的 2 秒噪音，所以受試者有 6 秒作反應。在確定受試者瞭解作答的形式與反應方式後，即開始施測。全部施測時間約為 120 分鐘。

本實驗設計採對抗平衡設計，即對每位受試者所呈現的表單次序不同。將 30 位受試者分為六組，每組呈現表單次序如表一所示：

表一 複本測驗之表單呈現先後次序

組別	表單呈現次序
第一組	1→2→3→4→5→6
第二組	2→3→4→5→6→1
第三組	3→4→5→6→1→2
第四組	4→5→6→1→2→3
第五組	5→6→1→2→3→4
第六組	6→1→2→3→4→5

3. 計分方式

將每張表單的 HP、LP 句子分開計算，每題 4 分，所以每張表單 HP 與 LP 句子得分範圍都是介於 0 至 100 分。再計算 HP 減去 LP 兩者的差距分數，所得數值以 HP-LP 差距分數表示。

(三)資料處理

1. 呈現 30 位受試者各在表單之平均數與標準差。

2. 應用 SPSS 統計軟體之皮爾遜相關 (Pearson product-moment correlation) 分析，計算受試者各表單兩兩相關之程度。

六、效度分析

(一)內容效度

本研究依測驗進行編製過程中所考量的幾項因素進行內容效度的分析，這些因素包括：1. 語音音節不重複的考量、2. 熟悉度的考量、3. 文意脈絡的考量、4. 句子的難度與合理性的考量、5. 預測度分析、6. 清晰度分析、7. 語音與字詞熟悉度分析等。

(二)構念效度

以 SPSS for Windows 8.0 版就六個表單進行統計資料處理。依 30 位受試者在各表單之 HP、LP 题目的得分進行因素分析。將各表單中 HP、LP 題目分為兩群，每一表單共計四個題目群，以斜交轉軸法 (oblique ration) 抽取因素。

(三)驗證性因素分析

以結構方程式 (structural equation model) 軟體 AMOS 3.6 版 (Arbuckle, 1997) 進行驗證性因素分析 (confirmatory factory)，檢驗測驗中六個表單之因素結構模式的適合度。

本研究以 30 位聽力正常年輕人之答題資料進行驗證性因素分析。研究者提出開放性模式 (open estimated model) 與限定性模式 (limited estimated model) 等兩種模式，分析向度包括三個部分：(1) 未標準化與標準化的參數估計值；(2) 開放性模式與限定性模式之 χ^2 值，以及 NFI (normed fit index)、CFI (comparative fit index)、NNFI (non-normed fit index)、GFI (goodness of fit index)、AGFI (adjusted goodness of fit index) 等比較不受樣本人數影響之指

標做為判斷模式適合度之指標；(3) 呈現 AIC (Akaike information criterion)、BCC (Browne-Cudeck criterion)、CAIC (consistent Akaike informatin criterion) 等求得模式複雜與簡約之平衡的簡約性指數。

結果與討論

一、預測度分析

表二呈現的刪題前後整體句子預測度的分布情況。結果顯示，刪題後比刪題前句子整體之預測度得分提高。經刪題後，有 84.32% 的句子預測度得分在 5 分以上。然而，刪題前只有 61.57% 的句子得分在 5 分以上。

表二 刪題前預測度得分分布統計表

預測度 分數	刪題前				刪題後			
	題數	百分比 (%)	累計 題數	累計百分比 (%)	題數	百分比 (%)	累計 題數	累計百分比 (%)
12	331	16.12	331	16.12	0	0.00	0	0.00
11	208	10.13	539	26.25	85	22.97	85	22.97
10	149	7.26	688	33.51	61	16.49	146	39.46
9	104	5.06	792	38.58	41	11.08	187	50.54
8	133	6.48	925	45.06	35	9.46	222	60.00
7	122	5.94	1047	51.00	34	9.19	256	69.19
6	100	4.88	1147	55.87	26	7.03	282	76.22
5	117	5.70	1264	61.57	30	8.11	312	84.32
4	123	5.99	1387	67.56	22	5.95	334	90.27
3	138	6.72	1525	74.28	22	5.95	356	96.22
2	122	5.94	1647	80.22	12	3.24	368	99.46
1	149	7.26	1796	87.48	2	0.54	370	100.00
0	257	12.52	2053	100.00	0	0.00	370	100.00
合計	2053	100.00	2053	100.00	370	100.00	370	100.00

二、清晰度分析

表三呈現的是刪題前後清晰度得分分布情形。就 HP 句子清晰度而言，刪題後所有題目均不是 19 分就是 20 分，比刪題前得分分布集

中，顯示刪題後 HP 句子的同質性提高。就 LP 句子清晰度而言，刪題後有 75.54% 的句子得分在 15 分以下，比刪題前的 71.02% 的句子得分略高。此外，就整體得分分布來看，刪題後 LP

句子的分布比刪題前集中，換言之，LP 句子的同質性提高了。再就 HP-LP 句子的得分差距之分布來看，刪題後有 73.37% 的句子在 5 分以

上，比刪題前的 47.03% 上還高。因此，刪題後整體句子之 HP-LP 得分和刪題前相較，達到了擴大 HP-LP 得分差距的目的。

表三 刪題前後清晰度得分分布統計表

清晰度分數	HP				LP				HP-LP			
	刪題前		刪題後		刪題前		刪題後		刪題前		刪題後	
	累計題數	累計百分比 (%)	累計題數	累計百分比 (%)	累計題數	累計百分比 (%)	累計題數	累計百分比 (%)	累計題數	累計百分比 (%)	累計題數	累計百分比 (%)
20	165	44.59	109	59.34	0	0.00	0	0.00	0	0.00	0	0.00
19	264	71.35	184	100.00	16	4.45	0	0.00	1	0.27	1	0.05
18	299	80.81	184	100.00	25	7.10	0	0.00	7	1.89	7	3.80
17	323	87.30	184	100.00	49	13.92	0	0.00	14	3.78	14	7.61
16	339	91.62	184	100.00	102	28.98	45	24.46	15	4.05	15	8.15
15	350	94.59	184	100.00	157	44.60	62	33.70	18	4.86	18	9.78
14	357	96.49	184	100.00	189	53.69	74	40.22	28	7.57	26	14.13
13	362	97.84	184	100.00	207	58.80	87	47.28	37	10.00	32	17.40
12	363	98.11	184	100.00	228	64.77	102	55.43	47	12.70	40	21.74
11	364	98.38	184	100.00	244	69.31	116	63.04	58	15.68	48	26.09
10	366	98.92	184	100.00	262	74.43	129	70.11	74	20.00	61	33.15
9	369	99.73	184	100.00	281	79.82	139	75.54	89	24.05	74	40.22
8	369	99.73	184	100.00	293	83.23	145	78.80	108	29.19	86	46.74
7	369	99.73	184	100.00	301	85.51	151	82.06	127	34.32	103	55.98
6	369	99.73	184	100.00	312	88.64	161	87.50	146	39.46	116	63.04
5	369	99.73	184	100.00	322	91.48	165	89.67	174	47.03	135	73.37
4	369	99.73	184	100.00	330	93.75	168	91.30	208	56.22	163	88.59
3	369	99.73	184	100.00	337	95.73	173	94.02	260	70.27	184	100.00
2	369	99.73	184	100.00	346	98.30	181	98.37	293	79.19	184	100.00
1	370	100.00	184	100.00	352	100.00	184	100.00	327	88.38	184	100.00
0	370	100.00	184	100.00	352	100.00	184	100.00	370	100.00	184	100.00
合計題數	370	100.00	184	100.00	352	100.00	184	100.00	370	100.00	184	100.00

三、語音與字詞熟悉度分析

(一) 分析結果

1. 音素分析

經清晰度分析後留下的 HP 與 LP 句子各 184

題。本研究依據每一個音節選取一個字的原則，刪除音節重複與句子不佳的 HP 與 LP 句子各 34 題，留下 HP、LP 句子共計 300 題。

2. 語音成分分析

語音成分分析共分為兩個部分，第一部分是分析整套測驗 300 個句子之整體句子語音成分。第二部分則是分析各表單 50 個句子的語音成分。以下先呈現 300 個句子語音成分分析的結果。

本測驗六個表單之 300 個句子的語音成分分布如表四所示。此外，本研究以教育部（民 89）常用語詞調查報告書之聲母與韻母出現頻率統計表與本研究中各句子語音成分做比較，結果顯示，句子的語音成分與教育部統計資料的百分比亦相當接近。

表四 300 個句子的語音成分分布統計表

發音方法	本研究		教育部
	數目	佔百分比(%)	佔百分比(%)
塞 音	975	39.52	30.48
擦 音	555	22.50	28.88
塞 擦 音	611	24.77	28.67
鼻 音	208	8.43	6.19
邊 音	118	4.78	5.78
聲母合計	2467	100.00	100.00
開 口 呼	1245	45.42	43.20
齊 齒 呼	783	28.57	29.79
合 口 呼	630	22.98	21.54
撮 口 呼	83	3.03	5.47
韻母合計	2741	100.00	100.00

註：教育部是指教育部（民89）常用語詞調查報告書之聲母與韻母出現頻率統計表

3. 熟悉度分析

熟悉度的分析包括前後兩次。第一次是在測驗編製前，依國語日報字典（民 85）所列 414 個音節結構延伸的所有字，請三位國小高年級同學判斷 (1) 是否學過這個字、(2) 這個字是否是他常用的字。合乎這兩項條件的國字則列入編製句子可能採用的關鍵字，經熟悉度考量後留下的關鍵字共計 1002 字。

編製成的句子，經句子難度與合理性的考量、預測度分析、清晰度分析、音素分析及語音成分分析後留下 150 個關鍵字，再進行第二次熟悉度分析。本研究依教育部（民 88）八十七年口語語料調查報告書之口語統計字頻表，計算 150 個關鍵字之出現頻次。口語統計字頻表共收字 3288 字，總頻次為 344209 次。本研

究 150 個關鍵字在口語中出現頻次的範圍介於 1 至 1300 次之間。

就熟悉度來看，Kalikow 等（1977）以桑代克字頻表（Thorndike-Lorge List）統計各關鍵字的熟悉度為依據。桑代克字頻表約收錄了 130 萬個字，其採樣來源是美語各報章雜誌及書籍上所收錄的字，該研究選取每百萬字出現 5 至 150 次做為關鍵字選取標準。本研究的熟悉度則考量包括兩部分，第一部分是先請三位國小高年級小朋友依自己的經驗勾選熟悉字。採用此步驟主要的原因是關鍵字的熟悉度會影響清晰度及預測度（Kalikow et al., 1977）。而隨著年齡增長，詞彙數量愈多，可能是造成兒童在國外版本 SPIN 表現不如成人的原因之一（Elliott, 1979）。本研究期望未來可以應用在

兒童的身上，因此在選取的過程與標準與國外略有不同。至於第二部分熟悉度分析主要是以教育部（民 88）口語語料調查報告書為依據，與 Kalikow 等（1977）採用書面文字之字頻有所不同。此外，教育部口語統計字頻表共收 3288 字，總頻次約 34 萬次，與國外收字總頻次不同。因此在選取題目時，過程不同、語系的差異、收字取材與總頻次的計算上等不同，都是本研究與國外研究在熟悉度考量上不盡相同之處。

本研究亦參考了國內兩項研究做為字詞熟悉度的比較。研究者以吳瑞屯、劉英茂（民 75）依書面文字統計的常用中文字出現次數統

計表為分析之依據，在常用中文字出現次數統計表中，本研究所有關鍵字出現的頻次範圍介於 2 至 6603 次之間。此外，將本研究關鍵字對照柯華葳、尹玫君（民 82）之國民小學常用字及生字難度研究報告，結果顯示本研究所有關鍵字均在國小學童常用字及難度範圍之內。

4. 各表單之語音分析

(1) 各表單清晰度之分析

各表單在分派後，HP 與 LP 句子清晰度之平均數與標準差如表五所示。就 HP 句子而言，所有表單的清晰度分數介於 19.48 至 19.64 之間。而就 LP 句子而言，所有表單的清晰度分數介於 10.44 至 12.36 之間。

表五 各表單清晰度平均數與標準差統計表

	表單1		表單2		表單3		表單4		表單5		表單6	
	平均數	標準差	平均數	標準差	平均數	標準差	平均數	標準差	平均數	標準差	平均數	標準差
HP	19.60	0.51	19.48	0.51	19.60	0.50	19.60	0.50	19.64	0.49	19.52	0.51
LP	11.84	4.38	11.64	4.35	10.44	5.09	10.44	4.67	11.12	4.76	12.36	4.38

本研究進一步以單因子變異數，考驗六個表單 HP 句子得分平均值有無顯著的差異。由

表六可知 $F=.343 (p>.05)$ ，顯示六個表單 HP 句子的清晰度沒有顯著差異。

表六 六個表單 HP 句子清晰度變異數分析摘要表

變異來源	SS	Df	MS	F
組 間	.433	5	.086	.343
組內誤差	36.400	144	.253	
總 和	36.833	149		

$P>.05$

至於六個表單之 LP 句子的清晰度變異數分析結果如表七所示。由表可得 $F=.744 (p>$

$.05)$ ，顯示六個表單 LP 句子之清晰度沒有顯著差異。

表七 六個表單 LP 句子清晰度變異數分析摘要表

變異來源	SS	Df	MS	F
組 間	76.053	5	15.211	.744
組內誤差	2945.840	144	20.457	
總 和	3021.893	149		

$P>.05$

(2)各表單 HP 句子預測度之分析 度之平均數與標準差。六個表單預測度平均數
表八呈現的是六個表單所有 HP 句子預測 介於 7.28 至 8.52 間。

表八 各表單 HP 句子預測度平均數與標準差統計表

	表單1	表單2	表單3	表單4	表單5	表單6
平均數	8.52	8.04	7.28	7.88	7.68	8.32
標準差	3.01	2.89	2.48	2.51	2.93	2.84

本研究進一步以單因子變異數分析六個表單 HP 句子預測度平均數，考驗六個表單 HP 句子預測度平均值有無顯著的差異。由表九可知 $F=.666 (p>.05)$ ，顯示六個表單 HP 句子之預測度沒有顯著差異。

表九 六個表單 HP 句子預測度變異數分析摘要表

變異來源	SS	Df	MS	F
組 間	24.913	5	4.983	.666
組內誤差	1077.760	144	7.484	
總 和	1102.673	149		

$P>.05$

(3)各表單之語音成分分析 現在表中。各表單句子之語音成分分布比率
前面提到句子的語音成分分析有兩部分， 大致與教育部（民 89）的資料相同。
第二部分所分析的是各表單 50 個句子，結果呈

表十 各表單 50 個句子之語音成分分布統計表

	表單一		表單二		表單三		表單四		表單五		表單六	
	數目	百分比	數目	百分比	數目	百分比	數目	百分比	數目	百分比	數目	百分比
塞 音	144	36.36%	152	39.58%	151	39.22%	148	39.36%	145	37.86%	167	42.39%
塞擦音	97	24.49%	92	23.96%	103	26.75%	95	25.27%	102	26.63%	93	23.60%
擦 音	90	22.73%	99	25.78%	77	20.00%	87	23.14%	92	24.02%	80	20.30%
鼻 音	37	9.34%	26	6.77%	33	8.57%	31	8.24%	31	8.09%	34	8.63%
邊 音	28	7.07%	15	3.90%	21	5.45%	15	3.99%	13	3.39%	20	5.08%
聲母合計	396	100%	384	100%	385	100%	376	100%	383	100%	394	100%
開口呼	210	46.46%	193	46.28%	182	43.13%	195	47.10%	203	47.43%	212	49.77%
齊齒呼	137	30.31%	111	26.63%	132	31.28%	123	29.71%	112	26.17%	105	24.65%
合口呼	94	20.80%	103	24.70%	94	22.27%	83	20.05%	101	23.60%	93	21.84%
撮口呼	11	2.43%	10	2.40%	14	3.32%	13	3.14%	12	2.80%	16	3.76%
韻母合計	452	100%	417	100%	422	100%	414	100%	428	100%	426	100%

(4)各表單關鍵字熟悉度之分析

就各表單關鍵字熟悉度來看，本研究以教育部（民 88）口語調查報告書之口語語料統計

字頻表為依據，統計各表單所有關鍵字平均出現頻次。表十一顯示出各表單關鍵字出現頻次介於 177.74 至 183.90 之間。

表十一 各表單關鍵字出現頻次分布統計表

	表單1	表單2	表單3	表單4	表單5	表單6
平均頻次	183.90	181.04	177.74	183.90	181.04	177.74

就各表單的音素分析、語音成分、清晰度、預測度及熟悉度等各項指標來看，各表單在各項指標的差異均不大，且符合日常生活之情境，顯示本測驗各表單各項指標之成分具有一致性。

四、信度分析與複本測驗的發展

表十二呈現的是 30 位受試者在六個表單的 HP 句子得分之平均數與標準差。各表單的 HP 句子得分之平均數介於 75.20 至 85.33 之間。而 LP 句子題目之平均得分介於 47.27 至 52.93 之間。再就 HP-LP 得分差距來看，表單的平均數介於 28.80 至 38.06 之間。

表十二 各表單 HP、LP、HP-LP 得分之平均數與標準差統計表

	HP		LP		HP-LP	
	平均數	標準差	平均數	標準差	平均數	標準差
表單1	81.47	11.30	49.60	14.43	31.87	12.84
表單2	75.20	15.67	48.00	12.95	27.20	13.39
表單3	85.33	13.83	47.27	12.48	38.06	10.27
表單4	85.33	13.18	49.20	11.89	34.13	10.02
表單5	80.40	12.66	51.60	10.98	28.80	10.73
表單6	83.20	14.50	52.93	12.89	30.27	12.02
平均	81.82		49.77		31.72	

至於六張表單其兩兩相關情形如表十三所示，六張表單不論在 HP、LP、HP-LP 得分差

距之所有相關係數均達到.05 的顯著水準，也由此可知六張表單彼此間相關程度很高。

表十三 各表單 HP、LP、HP-LP 得分之相關係數表

表單序號	表單1	表單2	表單3	表單4	表單5	表單6
表單1	1.000					
表單2	.850**	1.000				
表單3	.731**	.723**	1.000			
表單4	.817**	.750**	.826**	1.000		
表單5	.598**	.600**	.630**	.711**	1.000	
表單6	.884**	.812**	.816**	.876**	.702**	1.000

表十三 各表單 HP、LP、HP-LP 得分之相關係數表 (續)

表單序號	表單1	表單2	表單3	表單4	表單5	表單6
LP	表單1	1.000				
	表單2	.602**	1.000			
	表單3	.451**	.618**	1.000		
	表單4	.516**	.527**	.508**	1.000	
	表單5	.701**	.454**	.604**	.663**	1.000
	表單6	.566**	.643**	.428**	.649**	.448**
HP-LP	表單1	1.000				
	表單2	.346*	1.000			
	表單3	.414*	.478**	1.000		
	表單4	.467**	.509**	.358*	1.000	
	表單5	.469**	.462**	.392**	.363*	1.000
	表單6	.434**	.603**	.332*	.632**	.486**

註：*P<.05 **P<.01

五、效度分析

(一)內容效度

本研究在進行編製過程，在關鍵字的部分，做了音節不重複、字詞熟悉度等考量。而在句子部分亦進行了文意脈絡、句子的難度與合理性及語音成分等考量。最後再請專家審查。所有測驗的題目均先進行了預測度分析與清晰度分析，可說明本研究之測驗具有內容效度。

(二)構念效度

各表單的因素分析結果如表十四所示。表中之因素一代表的是 HP 句子題目，因素二代表的是 LP 句子題目。結果顯示：表單 1 中 HG12、HG11 屬於同一因素，LG11、LG12 屬於同一因素；表單 2 的 HG22、HG21 屬於同一因素，LG21、LG22 屬於同一因素。以此類推。

表十四 各表單 HP、LP 句子題目群與各因素之相關

表單序號	因素1 (HP)	因素2 (LP)	表單序號	因素1 (HP)	因素2 (LP)
表單1	HG12	.869	表單4	HG42	.948
	HG11	.842		HG41	.925
	LG11	.945		LG42	.945
	LG12	.932		LG41	.771
表單2	HG22	.914	表單5	HG51	.851
	HG21	.817		HG52	.835
	LG21	.966		LG52	.652
	LG22	.725		LG51	.960
表單3	HG32	.889	表單6	HG62	.942
	HG31	.854		HG61	.925
	LG31	.719		LG62	.876
	LG32	.830		LG61	.886

註：HG11 表示表單一第一組 HP 題目、HG12 表示表單一第二組 HP 題目、LG1 表示表單一第一組 LP 題目、LG12 表示表單一第二組 LP 題目，表單二、三、四、五、六依相同的方式呈現。

構念效度是指測驗能測量到理論上建構或特質的程度(陳英豪、吳裕益, 民 81)。在國外 SPIN 測驗中, 主要是想了解 HP 句子是否能代表聽覺閉鎖能力的表現, 而大多數研究皆指出 HP 比 LP 句子得分高(Bilger et al., 1984; Dirks, Bell, Robman & Kincaid, 1986; Elliott, 1983; Kalikow et al., 1977)。此外, Elliott 綜合文獻指出, 運用脈絡的能力會因是否有噪音的干擾而有不同的表現(Bilger et al, 1984)、句子脈絡字數的多寡會影響聽者預測的正確度(Dirk et al., 1986)、噪音會干擾聽者使用脈絡的能力(Pichra-Fuller et al., 1995), 上述這些證據說明聽覺閉鎖能力是一項特定的心理能力(Elliott, 1995)。除此之外, 尚沒有完全針對 SPIN 構念效度做探討的研究(Elliott, 1995)。本研究嘗試以因素分析方式, 抽取測驗題目中的因素。結果顯示六個表單之 HP 句子題目屬於同一因素, 而 LP 句子題目亦屬於同一因素, 並且 HP 比 LP 句子得分高。因此, 本研究結果或許可做為聽覺閉鎖能力是一種特定的心理能力的證據之一。

(三)複本測驗之驗證性因素分析

本研究以 30 位聽力正常年輕人之答題資料進行驗證性因素分析, 其驗證性因素分析的參數估計值如下: 就 HP 句子來看, 其開放性模式之標準化參數介於 .72 至 .96 之間; 而限定性估計模式之標準化參數介於 .80 至 .92 之間。再就 LP 句子來看, 開放性模式的標準化參數介於 .69 至 .78 之間; 限定性估計模式之標準化參數介於 .69 至 .81 之間。再就 HP-LP 分差距來看, 其開放性模式之標準化參數介於 .54 至 .79 之間; 限定性估計模式之標準化參數介於 .60 至 .74 之間。所有標準化的所有估計值均沒有負的誤差變異。

表十五呈現六個表單 HP、LP、HP-LP 之各種適合度指標。HP 句子在開放性模式的 $\chi^2=9.668$, $p=.378$; 限定性估計模式之 $\chi^2=$

19.607, $p=.143$ 。LP 句子在開放性模式的 $\chi^2=19.768$, $p=.019$; 而限定性估計模式之 $\chi^2=21.109$, $p=.099$ 。而 HP-LP 得分差距其開放性模式的 $\chi^2=6.045$, $p=.735$; 限定性估計模式之 $\chi^2=10.023$, $p=.761$ 。

然而 χ^2 值易受樣本人數影響, 只要樣本數夠大都會達到顯著水準。因此, 研究者參考 NFI (normed fit index)、CFI (comparative fit index)、NNFI (non-normed fit index)、GFI (goodness of fit index)、AGFI (adjusted goodness of fit index) 等比較不受樣本人數影響之指標做為判斷模式適合度指標。以上指標皆意指一個理論模式可以解釋實際資料的共變數百分比, 其值愈接近 1, 表示適合度越佳。一般而言, 大於 .90 就表示適合度極佳。就 HP 句子而言, 其開放性模式在 NFI、CFI、NNFI、GFI 均大於 .90; 而限定性模式在 CFI、NNFI 等指標大於 .90, NFI、GFI 亦接近 .90。就 LP 句子而言, 開放性模式在 CFI、NFI、GFI 接近 .90; 而限定性模式在 CFI、NNFI 大於 .90, GFI 接近 .90。就 HP-LP 得分差距而言, 其開放性模式在 NFI、CFI、NNFI、GFI 均大於 .90, AGFI 接近 .90; 而限定性模式在 CFI、NNFI、GFI 等指標大於 .90, NFI、AGFI 接近 .90。相較於虛無模式, 本研究之 HP、LP、HP-LP 之理論模式與觀察資料尚稱符合。此外, 本研究提出的兩種模式在各項適合度指標亦差異不大。

一般來說, 模式愈複雜, χ^2 值就愈小, 也就是適合度愈佳, 但是卻不具實用性。AIC (Akaike information criterion)、BCC (Browne-Cudeck criterion)、CAIC (consistent Akaike informatin criterion) 等指數是在求得模式複雜與簡約之平衡, 同時考量 χ^2 值與欲估計之參數個數, AIC 等指數越小表示越有可能找到適合的模式。表十六在 HP、LP、HP-LP 得分差距之簡約性指數均呈現, 虛無模式與本研究所提之兩個模式相較之下, 本研究所提之模式比較

表十五 六個表單 HP、LP、HP-LP 之各種適合度指標

		χ^2	df	p值	NFI	CFI	NNFI	GFI	AGFI
HP	虛無模式	187.106	15	0.000	0.000	0.000	0.000	0.000	0.000
	開放性模式	9.668	9	.378	.948	.996	.994	.901	.768
	限定性模式	19.607	14	.143	.895	.967	.965	.840	.761
LP	虛無模式	21.109	15	0.000	0.000	0.000	0.000	0.000	0.000
	開放性模式	19.768	9	.019	.803	.874	.790	.852	.654
	限定性模式	21.109	14	.099	.790	.917	.911	.843	.765
HP-LP	虛無模式	61.203	15	.000	.000	.000	.000	.000	.000
	開放性模式	6.045	9	.735	.901	1.000	1.107	.938	.855
	限定性模式	10.023	14	.761	.836	1.000	1.092	.903	.854

表十六 六個表單 HP、LP、HP-LP 之驗證性因素分析模式的簡約性指數

		AIC	BCC	CAIC
HP	虛無模式	199.106	202.924	213.513
	飽合模式	42.000	55.364	92.425
	開放性模式	33.668	41.304	62.482
	限定性模式	33.607	38.062	50.415
LP	虛無模式	112.308	116.199	126.788
	飽合模式	42.000	55.364	92.425
	開放性模式	43.768	51.405	72.415
	限定性模式	35.109	39.563	51.917
HP-LP	虛無模式	73.203	77.021	87.610
	飽合模式	42.000	55.364	92.425
	開放性模式	30.045	37.628	58.860
	限定性模式	24.023	28.477	40.831

能簡約地解釋受試者在六個表單彼此間互為複本，且本研究提出的兩種模式在各項簡約性指數差異不大。

國外針對 SPIN 測驗進行了三次複本測驗的發展 (Bilger, et al., 1984; Kalikow, et al, 1977; Morgan, et al., 1981)，這些研究大致採用各表單 HP、LP 句子及 HP-LP 得分差距之平均得分、標準差與兩兩相關等方式證明各表單間互為複本。本研究除了以皮爾遜相關分析，計算受試者在各表單得分之相關程度，亦嘗試以驗

證性因素分析檢驗複本測驗，除了可了解觀察資料與研究者所提的理論是否具有基本的適合度，尚採用比較不受樣本人數影響的 NFI、CFI、NNFI、GFI 及 AGFI 等適合度指標，以及 AIC、BCC 與 CAIC 等簡約性指數，結果顯示觀察資料大致符合本研究所提之理論模式，且本研究提出兩種模式在各項指標差異不大，顯示各表單之 HP 句子、LP 句子及 HP-LP 得分差距等分數均互為複本。

結論與建議

一、結論

(一)句子的編製

句子的編製從一開始就都是按句型規則編製 HP、LP 句子，並進行文意脈絡的考量、難度與合理性的考量。在關鍵字的選取上亦做了音節不重複與熟悉度的考量。且所有句子均經國小學童、教師與專家等人的判斷與審查，具有一定的合理性與難度。

(二)測驗的預測度

以 12 位高中職學歷以上的年輕人進行 HP 句子預測度分析，所有 HP 句子的預測度得分均介於 1 至 11 分間，顯示在噪音中聽取 HP 句子，不但可測量聽者使用文意脈絡訊息的能力，也能測量其在噪音使用語音音響訊息的能力。

(三)測驗的清晰度

清晰度分析結果呈現，HP 句子清晰度得分均介於為 20 分或 19 分、LP 句子清晰度則介於 1 至 16 分之間、HP-LP 得分差距介於 3 至 19 分之間。顯示 HP、LP 句子在清晰度上具有同質性，且兩者得分具有一定的差距。

(四)測驗的語音與字詞熟悉度分析

1. 本測驗的 HP 與 LP 句子題目共計 300 題，這 300 題包括了 150 個關鍵字，這 150 個字的音節結構均不重複，顯示本測驗的關鍵字具有語音不重複性。

2. 三百個句子之整體語音成分分析顯示，無論是 300 個句子的語音成分或是各表單 50 個句子的語音成分都與國語語音成份相近。

3. 就關鍵字的熟悉度來看，所有關鍵字的口語出現頻次介於 1 至 1300 次之間，在書面文字出現頻次則介於 2 至 6603 次之間，且所有的字均屬於國小學童認識的字與常用字範圍之內，顯示本測驗關鍵字具有一定的熟悉度。

4. 在測驗發展的六份表單中，每份表單在

預測度、清晰度、語音成分、熟悉度等各項成分均相近，顯示本測驗各表單各項成分具有一致性。

(五)測驗的信度分析與複本測驗的發展

六張表單在 HP 句子得分、LP 句子得分、HP-LP 得分差距等三種分數上，彼此相關程度很高，顯示六張表單具有複本信度，也說明了六張表單彼此互為複本。

(六)測驗的效度分析

1. 本測驗在編製過程做了關鍵字之音節不重複與熟悉度的考量，此外，也做了測驗的句子預測度與清晰度分析，且經專家審查，顯示本測驗具有一定的內容效度。

2. 六個表單分別進行因素分析後，都得到 HP 與 LP 兩個主要因素，顯示本測驗具有構念效度。

3. 以驗證性因素分析檢驗六個表單互為複本模式的適合度，結果顯示六個表單彼此之間互為複本的假設成立，且本研究提出兩種模式在各項指標差異不大。

二、建議

在國外，SPIN 測驗在 Kalikow 等 (1977) 發展以後，廣泛地被應用在許多研究及臨床工作上。而這些研究範圍涵蓋不同年齡層、不同聽力程度、不同群體在測驗信度與效度之驗證、相關能力的探討、應用情形等方面。或許這些研究主題可做為未來研究的參考，而研究者在研究過程中一些力有未逮的部分也可提供有興趣繼續探討這個主題的其它研究者參考，茲將建議陳述如下。

(一)在清晰度分析與樣本上的建議

本研究因研究者本身在時間、物力與財力有限，以及受試者參與長時間測驗之意願等因素，在清晰度測驗時，只選取 0 dB 及 -2 dB 兩個信噪比進行清晰度分析，未來研究可考慮增加信噪比數目做為清晰度分析的參考指標。此外，清晰度分析與複本測驗由於樣本人數較

少，使得統計結果在說服力上較為薄弱，未來研究可增加樣本人數，使測驗統計的數據更具代表性與說服力。

(二)測驗程序方面的建議

可考慮探討不同計分方式、傳輸體 (transducer)、作答方式、信噪比、噪音型式、句子預測度、句子音響特質等變項對測驗結果的影響，以提供未來應用在不同群體的一些訊息。

(三)不同受試群體

SPIN 測驗發展的目的之一，是期望能評估受試在真實生活的溝通能力，以了解其使用文意脈絡與語音音響訊息的能力。在國外已有許多研究探討過該測驗在不同群體之應用價值 (Agnew, 1997; Bilger, et al., 1984; Cokely & Humes, 1992; Elliott, et al., 1983; Elliott, 1995; Huchinson, 1989; Jerger, et al., 1989; Kalikow et al., 1977; Mayo, et al., 1997; Morgan, et al., 1981; Wild & Humes, 1990)，未來研究可探討聽力損失者、老年人、兒童、噪音中有聽取語音困難者及外語學習者在本測驗的表現。

(四)相關能力之探討的研究建議

Elliott (1995) 曾指出 SPIN 測驗所測量的是一種特定的聽覺閉鎖能力。未來或許可以探討聽覺閉鎖能力對工作記憶方面的影響、並進行說話理解模式的驗證。此外，也可藉由探討不同的語文能力、認知能力、注意力等相關因素對 SPIN 測驗表現的影響，以提供 SPIN 測驗在結果解釋時的一些額外訊息。

(五)在相關輔具的應用建議

國外版 SPIN 測驗有一些在相關輔具之運用研究成果 (Agnew, 1997; Kalikow et al., 1977; Wilde & Humes, 1990)，未來亦可考慮運用本測驗在助聽器、非聽覺輔助器、聽力保護用具等方面之探討。

參考書目

一、中文部分

- 吳瑞屯、劉英茂 (民 75)。中文字詞語音、語意屬性的研究。(行政院國家科學委員會研究計畫成果報告：NSC75-C301-H002-24)。臺北：行政院國家科學委員會。
- 柯華葳、尹玫君 (民 82)。國民小學常用字及生字難度研究報告。臺北：臺灣省國民學校教師研習會。
- 陳英豪、吳裕益 (民 81)。測驗與評量。高雄：復文。
- 教育部 (民 88)。八十七年口語語料調查報告書。臺北：教育部國語教育推行委員會。
- 教育部 (民 89)。八十七年常用語詞調查報告書。臺北：教育部國語教育推行委員會。
- 國語日報辭典 (民 85)。國語日報辭典。臺北：國語日報出版社。

二、英文部分

- Agnew, J. (1997). Binaural hearing for understanding speech in noise with hearing aids. *Seminars in Hearing, 18*, 355-373.
- Arbuckle, J. L. (1997). *Amos users' guide (Version 3.6)*. Chicago:Small Waters.
- Bilger, R. C., Nutzel, J. M., Rabinowitz, W. M., and Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech & Hearing Research, 27*, 32-84.
- Cokely, C. G., and Humes, L. E. (1992). Reliability of two measures of speech recognition in elderly people. *Journal of Speech & Hearing Research, 35*, 654-660.
- Dirk, D. D., Bell, T. S., Rossman, R. N., and Kincaid, G. E. (1986). Articulation index predictions of contextually dependent

- words. *The Journal of the Acoustical Society of America*, *80*, 82-92.
- Dubno, J. R., Dirk, D. D., and Morgan, D. E. (1984). Effect of age and mild hearing loss on speech recognition. *The Journal of the Acoustical Society of America*, *76*, 87-96.
- Elliott, L. L. (1979). Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. *The Journal of the Acoustical Society of America*, *66*, 651-653.
- Elliott, L. L., Lyons, L., and Busse, L. A. (1983). Speech understanding of older adults: A preliminary study. *The Journal of the Acoustical Society of America*, *73*, S24.
- Elliott, L. L. (1995). Verbal auditory closure and the speech perception in noise (SPIN) test. *Journal of Speech & Hearing Research*, *38*, 1363-1376.
- Hanks, W. D., and Johnson, G. D. (1998). HINT list equivalency using in older listener. *Journal of Speech, Language & Hearing Research*, *41*, 1335-1340.
- Hutcherson, R. W., Dirk, D. D., and Morgan (1979). Evaluation of the speech perception in noise (SPIN) test. *Otolaryngology Head Neck Surgery*, *87*, 239-245.
- Hutchinson, K. M. (1989). Influence of sentence context on speech perception in young and older adults. *Journal of Gerontology: Psychology Sciences*, *44*, 36-44.
- Jerger, J., Jerger S., Oliver T., and Pirozzolo F. (1989). Speech understanding in the elderly. *Ear and Hearing*, *10*, 79-89.
- Jerger, M. (1996). Phoneme awareness and the role of the educator. *Intervention in School & Clinic*, *32*, 5-13.
- Kalikow, D. N., Steven, K. N., and Elliott, L. L. (1977). Development of a test of speech intelligibility in noise testing sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America*, *61*, 1337-1351.
- Mayo, H. L., Florentine, M., and Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech & Hearing Research*, *40*, 686-693.
- Morgan, D. E., Kamm, C. A., and Velde, T. M. (1981). Form equivalence of the speech perception in noise (SPIN). *The Journal of the Acoustical Society of America*, *69*, 1791-1798.
- Pichora-Fuller, M. K., Schneider, B. A. and Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, *97*, 593-608.
- Plomp, P. (1986). A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *Journal of Speech & Hearing Research*, *29*, 146-154.
- Wilde, G., and Humes, L. E. (1990). Application of the articulation index to the speech recognition of normal and impaired listeners wearing hearing protection. *The Journal of the Acoustical Society of America*, *87*, 1192-1199.

Bulletin of Special Education 2002, 23, 121—140
National Taiwan Normal University, Taipei, Taiwan, R.O.C.

THE DEVELOPMENT OF MANDARIN SPEECH PERCEPTION IN NOISE TEST

Chih-Hao Tsai

Hsiao-Chuan Chen

Kaohsiung Municipal Min-Tsu elementary school

National Kaohsiung Normal University

ABSTRACT

The purpose of the study is to develop Mandarin Speech Perception In Noise test. The participants of this study were normal hearing adults aged between 20 to 30. Three groups of subjects were included to test the predictability (n=12), the intelligibility (n=31), and the equivalency (n=30) of the test lists.

The development of the test includes several procedures: predictability analysis, intelligibility analysis, word familiarity check, phonetic analysis, reliability and validity data analyses. The results of the study were listed as follows.

- 1.The predictability scores of this battery is from 1 to 11.
- 2.The intelligibility scores of the high-predictability sentences are 19 or 20, low-predictability sentences are between 1 to 16, and the HP-LP difference scores are between 3 to 19.
- 3.There are 150 different key words. The phonetic content of the sentences is representative of utterances used by speakers of Mandarin. Each form has a similar phonetic profile.
- 4.The word frequencies of word familiarity for key words are between 1 to 1300.
- 5.All of the list fit the equivalence model for high-predictability, low-predictability and HP-LP difference scores. It was found that the test has alternate-forms reliability.
- 6.The test has content and construct validity. Confirmatory factory analysis was conducted. It shows that six forms are equivalent forms.

Finally, there are some suggestions for the practical use and the future researches of the test.

Key words: Noise, Speech Perception, Intelligibility, Predictability, Word Familiarity