

成大台語檢定 ê 原理 kap 實務

◎蔣為文 國立成功大學台灣文學系副教授

1. 為啥物 ài 有台語能力檢定

當今 ê 社會分工真幼 mā 真注重專業證照。無論金融、資訊、電機、土木、語言等等 lóng 有相關 ê 檢定考試。行 òng 專業證照 ē-sái 講是現此時各行各業 ê 趨勢。

語言能力測驗 (language testing) tī 國外已經發展 kui-ā 十冬，而且有相當 ê 成就 ah。目前國內有常態辦理 ê 語言類證照考試包含有英語 ê TOEFL、TESOL、GRE、IELTS、TOEIC、GEPT 等，其他 ê 語種 koh 有日本語能力試驗、華語文能力測驗、客語能力認證、原住民族語語言能力認證等等。英語、日語等國際語言辦理檢定 ê 歷史 lóng khah 久，相對之下，台灣本土語言 ê 語言能力測驗制度是這 kui 年 tsiah 受 tiòh 重視 kap 開始發展。目前華語、客語、原住民族語已經分別有「國家華語測驗推動工作委員會」、「客家委員會」、「原住民族委員會」專責機構負責，而且開辦 kui-ā 冬 à。可惜，佔人口多數 ê 台語族群 suah 變成「弱勢 ê 多數」，除了 2002 年教育部臨時辦過 1 pái 台語認證之外，到 tann iáu m̄-bat 辦過全國性 ê 台語認證考試。今年 (2009) 後半年 13 縣市聯合委託成大台灣語文測驗中心辦理台語檢定，ē-sái 算是第一 pái 有經過學術研究 kap 預試過 ê 全國性台語能力檢定考試。

總是，犯勢有人會問講為啥物 ài 有台語能力檢定 (簡稱台檢)？啥人用會 tiòh？有啥好處？簡單講，台語能力檢定考試就親像「度針」kāng-khuán，醫生 ē-sái 用度針來量患者 ê 體溫 thang

判斷患者是毋是有發燒。台語能力檢定考試至少 ē-sái 運用 tī 下面 tsit 寡方面：

- 1) 了解台語教師本身 ê 台語能力是毋是有符合最 kē 要求。
- 2) 了解台語學習者 ê 台語能力是毋是有進步。
- 3) 做爲學校入學考、推甄入學 ê 時了解學生 ê 台語程度 ê 標準。
- 4) 做爲檢定台文系所學生畢業前 ê 台語能力 ê 標準。
- 5) 做爲公務人員、生理界、出版業等需要用 tiòh 台語人才 ê 時檢定 in ê 台語能力 ê 標準。
- 6) 了解台灣人民族母語流失 ê 速度。

2. 台語認證 / 檢定 ê 由來

教育部 uì 2001 年開始 kā 台語、客語 kap 原住民語列入國民中小學九年一貫課程大綱 ê 語文學習領域內底。因爲在職老師 ê 本土語言能力 khah 欠缺，爲 tiòh 補救 tsit-ê 欠點，教育部規劃用鄉土語言教學支援人員 ê 方式來鬥教本土語言 ê 課程。爲 tiòh 檢驗 in ê 語言能力，教育部 tī 2002 年 3 月辦理 1 pái 全國性 ê 台語認證考試。考試通過 koh 接受研習者 tsiah 有資格擔任台語科 ê 鄉土語言教學支援人員。這是台灣第一 pái 針對台語進行語言能力 ê 認證考試。因爲準備時間真短，所以並無預試 kap 相關 ê 信效度分析 tō 直接進行正式考試。

後來各縣市政府 mā 根據 ka-tī 縣市 ê 需要有辦過 kuí pái ê 地方性台語認證考試。全款，tsit 寡縣市所辦 ê 考試 mā 無經過預試 kap 統計分析。開始有用預試成績來做統計分析應該算是 uì 2006 年「全民台語能力檢定聯盟」tsiah 開始。

由成功大學台灣文學系台語研究室 kap 台南神學院等單位自發組成的「全民台語能力檢定聯盟」tī 2005 年 12 月 4 日第一 pái 召開籌備會議。經過 kuí pái 組織籌備會議 kap 認證規劃會議，路尾 tī 2006 年 9 月 23 日進行第一次試驗考試。Hit pái 預試分作台中（中山醫學大學）kap 台南（成功大學）2 個試場，參加考試的人數攏總 253 人。Hit pái 試考成果發表 tī 2006 年 12 月 2 日 tiàm 台南神學院辦理 ê 台語認證研討會。研討會當天正式成立「全民台語能力檢定聯盟」（簡稱「全民台檢」），koh 選出蔣為文擔任召集人^❶。

「全民台檢」籌備期間的行政事工由南神語言中心負責，測驗內容 kap 方式由「全民台檢規劃小組」負責。「全民台檢規劃小組」的主要成員有 6 位：蔣為文、何信翰、林裕凱、陳麗君、吳仁瑟 kap 丁鳳珍；同時由國立成功大學台灣文學系在職專班 kap 推廣教育台灣母語教學師資培訓碩士學分班超過 30 位的學員，協助試考相關細節，包括 1) 試題的研發及修整、2) 聽力測驗的試題錄音、3) 模擬考實施時的監考、4) 批改考卷以及 5) 其他雜務。所有 ê 工作人員 kap 試考考生 lóng 是義工，無領薪水。試務 ê 其他開銷，主要由鄭良偉教授 kap 海外 ê 台語運動者鄭良光先生贊助。Tī 第一 pái 試考了後，tī 2007 年 4 月 22 日 koh 進行第二 pái 試考，有 66 人參加。

Tng 學術界 kap 民間開始辦理台語能力檢定 ê 研發，教育部 mā 有認同 tsit-ê 研發工作 ê 重要。Tō 按呢「國語會」tī 2007 年 4 月 19 日第一 pái 召開「台灣閩南語言認證規劃會議」。經過 kuí

❶ 詳細請參閱丁鳳珍、陳麗君、吳仁瑟、林裕凱、何信翰、蔣為文 2006〈全民台語能力檢定測驗規劃 ê 現況〉收錄 tī 台語認證研討會論文集，台南，台南神學院。

pái ê 開會，路尾 tī 2007 年 11 月 21 日以台語字第 0960169012C 號令發布〈教育部臺灣閩南語語言能力認證作業要點〉。Hit-tsūn 教育部規劃用一冬做研發，suà--lài tī 2009 年正式開辦台語認證。教育部 ê 試題研發計畫後手由國立成功大學文學院台灣語文測驗中心得標，試務執行計畫由國立師範大學負責。

國立成功大學文學院台灣語文測驗中心 tī 2008 年底前已經辦理 2 pái 各 500 人 ê 預試，koh 已經完成認證方式 ê 規劃。無疑誤，2009 年 ê 認證經費 hōo 中國國民黨立法院黨團提案刪除，suah 無法度照進度辦理。知影預算 hōng 刪除了，真濟本土社團去立法院抗議兼去教育部陳情^②。事後行政院 kap 教育部 lóng 有表示會採取補救辦法。路尾由教育部補助有意願 ê 縣市政府自行辦理。因為成大台灣語文測驗中心目前是台灣唯一 ê 台語檢定單位，sóo-pái 各縣市 lóng tshuē 成大合作。第一 pái 正式台語檢定考試 tī 2009 年 11 月 14 號舉行，攏總有 793 名考生，主辦 ê 縣市包含台南縣市、嘉義縣市 kap 屏東縣。

國立成功大學台灣語文測驗中心 ê 組成成員除了有全民台檢 ê 人員之外，iáu koh 新增加一寡來自全國 ê 優秀人才^③。台灣語文測驗中心抱著



<http://ctlt.twl.ncku.edu.tw/>

- ② 會使參閱台灣羅馬字協會整理 ê 抗議活動新聞〈<http://www.tlh.org.tw/>〉，kap 何信翰 2009 〈台灣需要台語認證——台語認證簡介 kap 預算 hōo 立法院總 thài 記事〉《海翁台語文教學季刊》第 6 期，12-18 頁。
- ③ 詳細名單請參閱蔣為文主編 2009 《台語能力檢定實務導論》台南：亞細亞國際傳播社。

專業化、國際化 kap 本土化的遠景 kap 目標，來從事台語能力檢定的研發 kap 執行的 khang-khuè。咱測驗中心目前推出 ê 台語能力檢定是以中小學台語老師 kap 大學台文系所學生為主要對象 ê 教育用途 ê 語言檢定測驗。中心目前有 30 位顧問，18 位博士級的研發人員，以及 30 位大學以上的專任、兼任研究助理。團隊的成員主要是以成大為主，koh 招集全國各學術單位優秀的台語人才。Tsia-ê 優秀的台語人才，分別有各種語言教學、語言認證理論 kap 實務的經驗，包括台語、客語、原住民語、華語、英語、日語、德語、法語、俄語、越南語等等。Tī 專業研究領域頂面，包含一般語言學、語言教學、語言測驗、語音學、音韻學、句法學、構詞學、語意學、修辭學、語用學、社會語言學、心理語言學、方言學、文學、史學、統計學、教育學、自然科學、醫學等。博士級的研發人員攏有博士學位，koh 是 tī 大學 teh 教冊的助理教授（含）以上，in tī 語言認證相關的領域的理論論述方面有 khah 強的學術背景。除了學術界的專業研究員，koh 有 tī 第一線實際 teh 教學的國民中小學的老師 kap 台文作家參與在內。

3. 檢定測驗 ê 原理

咱 tsit 節會針對台語能力檢定 ê 原理 kap 規劃做簡單 ê 紹介 thang hōo 讀者有寡概念。代先，關係語言檢定 ê 用詞，到底「認證」、「檢定」、「測驗」有啥無全？

Tī 語言學界若 teh 討論按怎測量咱人 ê 語言表現能力（proficiency）ê 程度 ê 時，多數用語言「測驗」（language testing）ê 中性稱呼。也就是研究者單純針對受測者 ê 語言能力做程度 ê 客觀判斷，無摻加「通過」抑是「無通過」ê 主觀認定。

若是「認證」(certification)，通常是一个 khah 有權威 ê 機關，特別是政府單位，設定一定 ê 標準了，若考生有達到標準就算通過認證，而且會發證書 (certificate) hōo 考生。

「檢定」(qualification) kap 「認證」其實真類似，lóng 是有權威 ê 機關，設定一定 ê 標準了，對考生 ê 語言能力做評估。「認證」通常是 kan-tann 分「有通過」抑是「無通過」。「檢定」有可能分「有通過」kap「無通過」，mā 有可能附具體成績 kap 各級標準 hōo 需要者自行判斷。咱台語檢定就是採用 án-ne ê 做法。

除了「認證」、「檢定」、「測驗」ê 用法，iáu koh 有評估 (evaluation) kap 評量 (assessment) 等 ê 用詞，通常評估 kap 評量是用 tī 語言教學 ê 課程 lìn 對學生 ê 學習成果做檢驗。

語言測驗是毋是可靠、有效，就 ài 看伊 ê 「信度」(reliability) kap 「效度」(validity)。所謂 ê 信度，是指 tī kāng-khuán iáh 是類似 ê 條件下重複試驗 ē-sái 得 tiòh 一致 iáh 穩定 ê 結果。效度是指檢定工具 ē-sái 準確、有效 ê 量出考生 ê 語言能力 ê 程度。根據 2008 年 11 月 500 人試考結果，成大台語能力檢定 ê 信度 Cronbach's α 值達到 0.873，算是真 kuân ê 數字；效度上 kuân 達到 $\gamma_{pb}=0.34$ ， $\gamma_{bi}=0.54$ ；預測準確度達 0.68。

語言測驗 ē-sái 分做「常模參照測驗」(norm-referenced test) kap 「標準參照測驗」(criterion-referenced test)。Beh 採用佗 1 種 ê 測驗方式，就 ài 看測驗 ê 目的 kap 對象。

咱成大台語檢定是以教育用途為目的，主要是 beh 測看國內台語老師 kap 學生 ê 台語程度到佗位。學生主要設定 tī 高中、

大學 kap 研究所等級。照講台語是台灣人 ê 母語，台灣人應該有 tsát-pak ê 台語能力，無需要做能力檢定。照講台語檢定應該 hām TOEFL kāng-khuán 是以外國人做測驗 ê 對象，毋過因為台語流失 siūnn 嚴重，suah tiòh 以國內 ê 台灣人為測驗 ê 主要對象。

咱成大台語檢定採用「標準參照測驗」，主要原因是：1) 咱需要 1 套標準 thang 知影考生 ê 台語程度是落 tī 啥位置，2) 參加台檢考試 ê 人數無夠大，in ê 台語能力 mā 無一定有常態分布，所以無適合用常模參照。

設計語言測驗就親像設計 1 支「度針」來度體溫 kāng-khuán。度針是以水 ê 冰點定義做 0 度，以滾點（沸點）做 100 度，冰點 kap 滾點之間分做 100 kuéh，按呢每 1 kuéh 就號做攝氏溫度 1 度。台語檢定以台語是第一語言 ê 人做「上 kuân 級」，以完全 bē-hiáu 台語 ê 人做 0 級；兩者當中 tsiah 根據需要 kap 指標來分級數。這部分台檢是採用歐盟 ê 標準，kā 台語分做 6 个（若包含 0 級，就算 7 个）級數。

語言能力級數決定好勢了，有 2 種考試方式。第一種是各級數分開考試，譬如初級 kap 中級分 2 pái 舉行。第二種是各級數做伙考 tsiah 根據考試成績來分考生 ê 所屬級數。這 2 種做法各有優點 kap 欠點。台檢就是採用第二種 ê 做法。

測驗用 ê 題目會根據語言能力級數來分級。因為台檢是採用「分級，毋過各級數做伙考」，所以 1 份考試單內底會包含所有級數 ê 題目。基本上，考生 ê 考試結果若分數愈 kuân，就表示伊 ê 語言能力愈好。另外，為 tiòh tshiau-tshik 各級數所對應 ê 成績分數，通常會設計「樣本考生」來試考看 māi。樣本考生是根

據語言能力分級指標來揀選，in 包含所有 ê 級數。Suà--lâi 研究者會 kā 試考用 ê 各級題目 thêh hōo 各級 ê 樣本考生試考。試考了 ê 成績 tsiah koh thêh 來檢討試題 kap 樣本考生 ê 分級。經過 kuí pái ê 試考 kap 交互檢討，路尾 tsiah koh 利用樣本考生試考 ê 成績來做 regression ê 分析，推測出各語言級數對應 ê 分數。

4. 分級標準

各種語言檢定所採用 ê 分級標準無一定 kâng-khuán。台檢 ê 分級標準是採用「歐洲理事會」(Council of Europe ④) tī 1996 年公布 ê 語言能力分級 (Common European Framework of Reference for Languages: Learning, teaching, assessment, 簡稱 CEF) 標準，分做六級。讀者 ài 注意，tī CEF ê 分級，A 算是 khah 初級 ê，C 是 khah kuân 級 ê (能力 khah 好 ê)。Tse kap 咱一般認為 A 是 khah 好 ê tú-hó 倒反。

圖表 1. 台語能力檢定 ê 語言能力分級標準

分級標準	分級標準	標準說明
Basic User	A1 Breakthrough 基礎級	Ē-tàng 瞭解 koh 使用簡單 ê 語詞來應用、表達 tī 日常 ê 基本需要，koh ē-tàng 紹介 ka-kī kap 別人 ê 背景資料。若是對方講話速度 khah 慢、用詞發音清楚，koh 願意提供適當 ê 協助，ē-tàng 做簡單 ê 交流。
	A2 Waystage 初級	Ē-tàng 理解大部分 kap ka-kī 有相關 ê 語句 kap 常用語 (譬論：Ka-kī kap 家庭 ê 基本資料、買物件、當地 ê 地理環境、khang-khuè)。Tī tshiāng-tsāi 做 ê 簡單事務，ē-tàng kap 人溝通 kiau 簡單、清楚 kap 人交換資訊。Ē-tàng kā ka-kī ê 背景、周圍 ê 環境、事務狀況等等，做簡單 ê 說明。

④ Council of Europe 網址：<http://www.coe.int>。中文版請參閱莊永山 2007《歐洲共同語文參考架構》高雄：多媒體英語學會。

分級標準	分級標準	標準說明
Independent User	B1 Threshold 中級	Tī khang-khuè 頂頭、學校、îng-siān 等等場合，tīng-tiòh khah 熟 ê 事務 ê 時，收 tiòh 清楚 koh 標準 ê 訊息了，ē-tàng 瞭解伊 ê 重點。Tī 目標語言地區旅遊 ê 時，ē-tàng 應付大部分會出現 ê 狀況。對 ka-kī 熟似 á-sī 有興趣 ê 主題，ē-tàng 寫簡單 ê 短文。Ē-tàng 講出 ka-kī ê 經驗、事件、夢想、ng 望 kap 志向，對 ka-kī ê 看法 kiau 計畫 ē-hiáu 做簡單 ê 解釋 kap 說明。
	B2 Vantage 中高級	Tui 具體 kap 抽象主題 ê 複雜文章，lóng ē-tàng 瞭解伊 ê 重點，包含 ē-tàng 用個人專業領域來 tui 文章做技術性 ê 討論。Ē-tàng 真順 bē gāi-giòh kap 母語使用者主動對話。Ē-tàng 針對 bē-tsió ê 主題寫一份完整詳細 ê 文章，koh ē-tàng 針對 tàk 項議題做重點 ê 優、欠點說明。
Proficient User	C1 Effective Operational Proficiency 高級	Ē-tàng 瞭解內容範圍闊，難度 kuân ê 長篇文章 koh ùi 字面知影伊另外 ê 語意。Ē-tàng 真 kùt-liu、自然表達 ka-kī，koh bē bó 話 thang 講。Tī 社交、學術 kap 專業 ê 目的，ē-tàng 有 tshun-kiu，有效來使用語言，koh ē-tàng 清楚詳細架構完整來寫複雜 ê 議題。
	C2 Mastery 專業級	Ùi 實際所聽 tiòh、讀 tiòh ê 訊息，ē-tàng 真輕鬆 tō 瞭解。Ē-tàng ùi bò-kāng ê 口頭 kap 書面資料，做重點說明。Mā ē-sái 重新建構內容內底 ê 論點，做相連 suà ê 表達。甚至 tī koh khah 複雜 ê 情形，mā ē-tàng 真自然、kùt-liu koh 正確講出另外 ê 語意 kiau 意涵。

5. 基本詞頻、句法點 kap 用字

語言測驗 ê 時，考生 ê 語言能力是由考試結果 ê 分數來做指標。考生 ê 得分是根據伊回答 guā-tsē 題目來決定。因為考卷包含各級難易度 ê 題目，所以回答 tiòh ê 比例愈 kuân 表示考生 ê 程度愈好。若是按呢，beh 按怎決定題目 ê 難易度？台檢是利用基本詞頻、句法點 kap 實際試考來決定題目 ê 級數。

所謂 ê 詞頻是指 hit-ê 語詞出現 ê 頻率，也就是 tsiáp 出現 iáh 無 tsiáp 出現 ê 情形。通常，咱 ê 假設是愈少出現 ê 語詞（詞頻低）in ê 難度就愈 kuân。雖罔是按呢，詞頻並無百分百反應難度，有時 mā 有例外。Tse 是因為詞頻會因為使用者、使用場合、專業領域等因素造成頻率 ê 無全。所以語詞 ê 難易度咱 iáu 會以實證 ê 試考結果來調整。句法 ê 難易 mā 會影響 tiòh 題目 ê 級數。台檢團隊 ê 做法是先依照專業經驗判斷台語句法 ê 難易，suà--lâi tsiah koh 根據試考 ê 結果做調整。

台語能力檢定 ê 用字原則是：第一，ài 用漢字 ê 部分一律使用教育部公布 ê 推薦用字。第二，有羅馬字 ê 部分一律使用教育部公布 ê 臺羅拼音正式版。第三，以「漢羅」（也就是漢字 + 羅馬字）ê 方式呈現，羅馬字 ê 比率大約佔 5%-10%。Tang 時用漢字 iáh 是羅馬字 tō 看語句 ê 情境來判斷，以 ē-tàng hōo 語句簡單明瞭為原則。

6. 現行考試科目、題型 kap 配分

咱 2009 年正式推出 ê 台語能力檢定考試考試科目、題型內容、考試時間 kap 配分分別列 tī 圖表 2。詳細的介紹請參閱本期其他文章或者《台語能力檢定實務導論》。

圖表 2. 考試科目、題型內容、考試時間 kap 配分

考試科目	考試時間	分數配分	
閱讀測驗 (a) 詞彙 kap 語法測驗 (36 題) 1. 詞彙測驗 (24 題) 2. 語法測驗 (12 題) (b) 閱讀理解 (24 題)	70分鐘	(a) 108 分	(b) 72 分
		小計 180 分	

考試科目	考試時間	分數配分	
聽力測驗 (a) 對話選擇題 (24 題) (b) 演說選擇題 (16 題)	40分鐘	對話 72 分	演說 48 分
		小計 120 分	
聽寫測驗 語詞聽寫 (40 題雙音節以上的羅馬字書寫)	20分鐘	小計 80 分	
口語測驗 (6 題) (a) 看圖講古 (2 題): (每題準備時間 30 秒, 作答時間 1 分鐘) (b) 朗讀測驗 (2 題): (每題 300 字以內, 每題時間準備 30 秒, 作答時間 2 分鐘) (c) 口語表達 (2 題): (每題準備時間 1 分鐘, 作答時間 2 分鐘)	30分鐘	每題 20 分	
		小計 120 分	
合 計	160分鐘	500 分	

7. 考試成績 kap 語言能力級數對應

成大 tī 2008 年 11 月辦理第二 pái 預試, 試考 ê 成績列 tī 圖表 3。

圖表 3. 2008 年 11 月預試考生 ê 總成績 kap 各題型 ê 平均得分

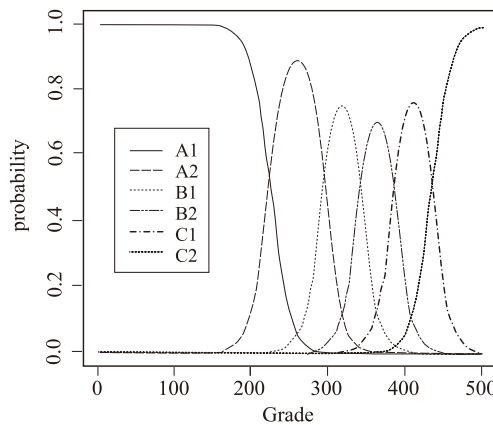
考試科目	有效人數	得分平均	得分標準差	分數分配
一、閱讀測驗	378	109.25	24.81	180
詞彙及語法測驗	378	69.50	16.55	108
閱讀理解	378	39.75	12.71	72
二、聽力	380	79.29	17.89	120
對話選擇題	380	52.29	11.15	72
演說選擇題	380	27.01	9.45	48
三、聽寫	380	54.30	23.27	80

考試科目	有效人數	得分平均	得分標準差	分數分配
四、口語測驗	365	82.89	18.80	120
看圖講古	365	30.25	6.45	40
朗讀測驗	365	25.00	7.59	40
口語表達	365	27.64	7.43	40
總分	363	327.99	61.30	500

台檢採用歐盟 CEF ê 標準 kā 台語能力分做 6 級。考生 ê 成績分數 kap 伊對應 ê 語言能力級數 beh 按怎決定？台檢是利用 160 名 ê 樣本考生試考 ê 成績做解釋變數，透過「順序羅宜斯迴歸」（Ordinal Logistic Regression）模型計算各級 ê 分數。

圖表 4 是樣本考生所建構出 ê 順序羅宜斯迴歸圖，y 軸代表機率，x 軸代表總成績。根據圖表，分數越低，考生是 A1 級 ê 機率愈 kuân；分數愈 kuân，考生是 C2 級 ê 機率愈 kuân。圖 lín lóng 總有 6 个山谷形 ê 線，線 kap 線 ê 交點就是各級 ê 界線。這 6 个山谷 lóng 總有 5 个交點，分別是 220、290、340、380、430。Uì 倒算過來第二個山谷是 A2 出現 ê 機率，伊 kap 第一 hām 第三 ê 山谷 ê 交點分別是 220 kap 290。因為這段範圍之內 A2 ê 機率上大，咱 tō 講 220 kap 290 分之間屬 A2 級能力。

圖表 4. 樣本考生所建構出 ê 順序羅宜斯迴歸圖



圖表 5. 台語語言能力級數 kap 考試成績對應

分級標準	考生成績總分* (成績滿分 500 分)	成大建議
A1 Breakthrough 基礎級	150 < 總分 ≤ 220	建議來台留學 ê 外籍生 (特別是 tī 南台灣留學者) ài 達 tsit 級 thang 應付生活 ê 需要。
A2 Waystage 初級	220 < 總分 ≤ 290	建議來台留學而且就讀台文相關系所 ê 外籍生 ài 達 tsit 級 thang 應付生活 kap 學業 ê 需要。
B1 Threshold 中級	290 < 總分 ≤ 340	平時有法度用口語參人溝通, m̄-koh 無受過讀、寫訓練者, ài 達 tsit 級。
B2 Vantage 中高級	340 < 總分 ≤ 380	大學修過台語相關課程 4 學分 iáh 是參加研習 72 小時以上者、國中小台語老師 ài 達 tsit 級。
C1 Effective Operational Proficiency 高級	380 < 總分 ≤ 430	台灣語文學系 iáh 台語文教學學程 ê 畢業生 ài 達 tsit 級。
C2 Mastery 專業級	430 < 總分 ≤ 500	從事台語師資培訓 ê 種子師資 ài 達 tsit 級。

* 聽、說、讀、寫四科 bē-sái 有欠考 iáh 零分

8. 話尾

成功大學台灣語文測驗中心是全台灣第一個專門針對台語能力檢定做研發 kap 執行 ê 學術單位。台語檢定 ē-sái tī 2009 年 11 月順利辦理正式考試, ài 感謝真濟人 ê 鬥相共 kap 參與。向望透過台語檢定 ê 推行 ē-sái 提升台語教學 ê 品質 thang 進一步挽回台語 ê 活力。

台語能力測驗發展現況

蔣為文

國立成功大學台灣語文測驗中心

摘要

台語能力測驗是為tiòh教育目的所開發 ê 新 ê 語言測驗系統。成大辦理台語能力測驗 ê 正式名稱號做全民台語認證，伊是經過kuí擺 ê 預試了tsiah正式開辦 ê 考試。本文以2008年11月所做 ê 預試資料做信、效度分析。Tī全民台語認證，考生會根據in得tiòh ê 分數來決定in ê 台語能力級數。成績kap能力 ê 對照關係是根據預試ê 160名樣本考生做模型tsiah推算出來。根據預試結果，預測ê正確率達到0.68。

關鍵詞：台語、GTPT、CTLT、能力測驗、語言測驗

收件日期2010.05.17/接受日期2010.06.23

The Development of the Taiwanese Proficiency Test

Wi-vun Taiffalo CHIUNG
Center for Taiwanese Languages Testing
National Cheng Kung University

Abstract

The Taiwanese Proficiency Test (TPT) is a newly developed language testing scheme for educational purposes in Taiwan. Several preliminary tests were conducted and the latest one was held in November of 2008. In the test, the test-takers were assigned a language level according to the scores they received. The relationship between language levels and scores is calculated based on the statistical results of 160 sampled test-takers by using ordinal logistic regression. Its accuracy for predicting test-takers' language levels reaches as high as 0.68.

Keywords: Taiwanese, TPT, CTLT, proficiency, language test

1. Introduction

Taiwan is a multilingual and multi-ethnic society. Traditionally, it is divided into four primary ethnic groups. Because nation-wide linguistic census have not been conducted in recent decades, no accurate ethnolinguistic demographics are available. However, according to frequently cited data, the speakers of each ethnic group were estimated as follows: indigenous peoples (1.7%), Hakka (12%), Taigi or Taiwanese (73.3%), and Mainlanders ¹ (13%) (Huang, 1993:21). In addition, as mixed-cultural marriages have become more common in the globalization era and Taiwan being no exception, according to the statistics of Taiwan's Ministry of Interior (2009), there were a total of 414,699 foreign spouses in Taiwan by January 2009 and they constitute the fifth ethnic group in Taiwan.

In addition to being a multi-ethnic society, Taiwan has been colonized by several foreign regimes since the seventeenth century. Two centuries later, the sovereignty of Taiwan was transferred from China to Japan in 1895 as a consequence of the Sino-Japanese War. At the end of the World War II, Japanese forces surrendered to the Allied Forces. Chiang Kai-shek, the leader of the Chinese Nationalist (KMT ² or Kuomintang) took over Taiwan on behalf of the Allied Powers under General Order No.1 of September 2, 1945 (Peng, 1995:60-61). At the time, Chiang Kai-shek was fighting against the Chinese Communist Party in Mainland China. In 1949, Chiang's troops were completely defeated and then pursued by the Chinese Communists. At that time, Taiwan's national status was supposed to be dealt with by a peace treaty among the fighting nations. However, because of Chiang's defeat in China, Chiang decided

¹ Mainly the immigrants came to Taiwan with the Chiang Kai-shek's KMT regime around 1945.

² KMT was the ruling party in Taiwan since 1945 until 2000, in which year Chen Shui-bian, the presidential candidate of opposition party Democratic Progressive Party was elected the new president. Thereafter, the KMT won the presidential election again in 2008, and has become the ruling party again since 2008.

to occupy Taiwan as a base from which he would fight back and retake Mainland China (Kerr, 1992; Ong, 1993; Peng, 1995; Su, 1980). Consequently, Chiang's political regime the Republic of China (R.O.C) was relocated and resurrected in Taiwan and has remained there since 1949.

The National Language Policy ³, or monolingual policy, was adopted both during the Japanese and the KMT occupations of Taiwan. In the case of the KMT's monolingual policy, the Taiwanese people were not allowed to speak their vernaculars in school and in public. Moreover, they were forced to learn Mandarin Chinese and to identify themselves as Chinese through the national education system (Cheng, 1996; Tiun, 1996). As Hsiao (1997:307) has pointed out, "the usage of Mandarin as a national language became a testimony of the Chineseness of the KMT state." Consequently, researches such as Chan (1994) and Young (1988) have revealed that a language shift toward Mandarin is in progress. Huang (1993:160) goes so far as to suggest that the indigenous languages of Taiwan are all endangered.

Mother tongue education was not implemented nation-wide until 2001, the year after the KMT lost the presidential election for the first time in Taiwan. Since then, all elementary school classes are required to have a class called "local language", lasting 40 minutes, once a week in school. The schools may choose which local language to teach in accordance with the demands of the student population order to ensure that all local language teachers have a certain level of local language proficiency, proficiency tests in local languages had been planned and administered. Three language tests were prepared. The indigenous languages test has been planned and executed by the Council of Indigenous Peoples, Executive Yuan, since 2001. The Hakka language test was prepared by the Council for Hakka Affairs, Executive Yuan, and the first official test was conducted in 2005. As for Tai-gi or the Taiwanese language, since there is no special council for Tai-gi speakers, the task for Taiwanese proficiency test was

³ For details, see Huang 1993.

taken over by the National Languages Committee (NLC) of the Ministry of Education (MOE).

2. Historical background of Taiwanese language testing

Although mother tongue education has been officially included in elementary schools starting from the 2001 academic year, there was no Taiwanese proficiency test held for teachers to teach Taiwanese in classrooms. Because most elementary teachers were neither fluent in spoken Taiwanese nor educated in written Taiwanese as a consequence of monolingual policy, they faced problems in teaching Taiwanese. To overcome this predicament, several proposals were put forward. Among the proposals, a policy for teaching assistants was adopted. That is, a teaching assistant who does not have a teacher's certificate but has a good knowledge of the Taiwanese language will be in charge of language teaching along with the classroom teacher. To find linguistically qualified assistants, a hastily-held proficiency test in Taiwanese took place nationally in March of 2002. The persons who passed the test had to take a 36-hour training course before he or she can teach in the classroom. Because the test was given in a rush, no preliminary trials were conducted. In other words, neither reliability nor validity was analyzed. After that, Taiwan's Ministry of Education (MOE) gave authority to local governments to hold language tests in local languages. As a result, some local governments, such as Tainan county, Tainan city and Kaohsiung city, held their versions of the Taiwanese proficiency tests. However, due limited budgets and professional resources, the tests were not well planned and conducted. The criteria of proficiency varied from place to place.

A professional Taiwanese test was not well planned until 2006, when the General Taiwanese Proficiency Test League (GTPTL) was founded. GTPTL was first convened at Tainan Theological College and Seminary on December 4, 2005. The league consisted of academic institutions and Taiwanese language associations, such as Tainan Theological College and Seminary, Department

of Taiwanese Literature of National Cheng Kung University, Department of Taiwanese Languages of Chung Shan Medical University, and Taiwanese Romanization Association (Teng, 2006). After several months of preparations, the first preliminary test was conducted with a total of 253 subjects on September 23, 2006. Its statistical results were presented at the first conference on Taiwanese proficiency test held by Tainan Theological College and Seminary on December 2, 2006 (Chiang, 2006). Thereafter, a second pilot test was arranged with a total of 66 subjects on April 2007. An official Taiwanese proficiency test was proposed by the GTPTL to be conducted in 2008. However, the test was postponed because major members of GTPTL were commissioned by the Ministry of Education to work on a research project to develop a new national level test in Taiwanese.

It turned out that while the GTPTL was working on the Taiwanese test, the National Languages Committee (NLC) of the MOE decided to develop a Taiwanese test, too. Its first meeting on the planning of the test was convened on April 19, 2007, and finally an Operating Guideline for Language Proficiency Test in Taiwanese Southern Min was promulgated on November 21, 2007. According to the resolution of the meetings, the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (Thereafter, CEF) was adopted as a guideline for the proposed Taiwanese test. Preliminary tests were scheduled for 2008 and an official nationwide test in 2009.

In accordance with results of public biddings, National Cheng Kung University was given charge of the research project for 2008, and National Taiwan Normal University administrative duty of executing the test in 2009. The members of GTPTL became the major constituents of the research team. During the research periods, two preliminary tests were conducted on August 23 and November 29, 2008, respectively. Each test contained 500 subjects. Although the preliminary tests were done well and the official test was expected to be conducted in August 2009, the budget for administrative affairs was suddenly canceled by the KMT legislators in early January 2009. Consequently, the official nation-wide test that was originally scheduled to take place in 2009 was forced to

be terminated. In response, more than twenty grassroots organizations protested against the KMT legislators on February 27, 2009.⁴ Under the pressure of the grassroots organizations, the MOE promised to subsidize local governments as a remedy for canceling the budget for the test. In other words, Taiwanese tests will be conducted by counties/cities, rather than by the central government. According to the meeting convened by MOE's Department of Elementary Education on May 26, 2009, there were 13 counties/cities willing to hold the Taiwanese test. All the counties/cities agreed to appoint the research team of National Cheng Kung University (NCKU) as the planner to carry out the test. The first official test was co-organized by Tainan City, Tainan County, Chiayi City, Chiayi County, and Pingtung County. It was scheduled for November 14, 2009, and 793 test-takers were registered.⁵

3. Formats of the current Taiwanese proficiency test

The Taiwanese Proficiency Test (TPT) described in this paper was designed by the Center for Taiwanese Languages Testing (CTLT), National Cheng Kung University. The initial purpose for developing TPT was to measure Taiwanese language proficiency of elementary school teachers and language teaching assistants to ensure that they fulfill the requirement of minimal language ability. Later it was expanded to include college students majoring in Taiwanese and to all members of the public (adults only). The current format, which was adjusted and revised based on 4 preliminary tests, was carried out for the first official test in November 2009. Table 1 presents the format of the current Taiwanese proficiency test designed by the CTLT.

⁴ For more information on the demonstration, visit <<http://www.TLH.org.tw>>

⁵ For details on the test, visit <<http://ctl.t.wl.ncku.edu.tw>>

Table 1. Format of the Taiwanese Proficiency Test by CTLT

Sections	Time	Scores	
Reading (a) Vocabulary and grammar (36 questions) 1. Vocabulary (24 questions) 2. Grammar (12 questions) (b) Reading comprehension (24 questions)	70 mins.	(a) 108	(b) 72
		subtotal: 180	
Listening (a) Conversations (24 questions) (b) Talks and lectures (16 questions)	40 mins.	(a) 72	(b) 48
		subtotal: 120	
Dictation Word dictation in Tai-lo Pheng-im (40 words)	20 mins.	80	
Speaking (a) Using picture prompts for storytelling (2 questions) (b) Oral readings (2 questions) (c) Oral expressions (2 questions)	30 mins.	20 for each subtotal: 120	
Total	160 mins.	500	

Generally speaking, norm-referenced and criterion-referenced tests are the major approaches in language testing (McNamara, 2000, pp.62). Criterion-referenced measurement was adopted by the CTLT for the TPT. The major reasons are as follows: 1) The TPT was initially designed for examining Taiwanese teacher's Taiwanese language proficiency level. The criteria for language levels were set in advance. The purpose of the TPT is to locate teachers' standardized language level in accordance with the criteria, rather than finding a teacher's relative level among all the teachers. 2) The number of test-takers was not expected to be high.

The TPT consists of 4 sections: reading, listening, dictation, and speaking tests. The total score is 500 points. The TPT divides Taiwanese language

proficiency into 6 levels in accordance with the CEF criteria, that is, A1, A2, B1, B2, C1, and C2. Test items of an individual TPT test comprise all six language levels. In other words, a test-taker does not have to take six tests, from the most basic to the most advanced level to locate his/her level. Instead, a test-taker needs to take only one test and s/he will be assigned a language proficiency level depending on the score s/he gets, as shown in TABLE 9. The main reasons for not creating tests in different proficiency levels are: 1) the budget and resources for the TPT are not sufficient to hold 6 individual tests of different language levels in a year. It can only hold tests once or twice a year under the current conditions. 2) The test needs to be done as quickly and efficiently as possible to find out the Taiwanese teachers' language ability. It is an economic way to include all test-takers of different language levels in one test.

The first issue that needed to be solved was the selection of a Taiwanese writing system that will be used consistently throughout the test. Taiwanese writing systems are either in Han characters, Roman alphabet or a mixed system combining the two systems. Currently, the dominant writing system is called Han-lo, or literarily Han characters plus Roman scripts. However, different users may have different opinions on choosing the Han characters or the Romanization schemes (Chiung 2001). To standardize Romanization, Tai-lo Pheng-im, a Romanization scheme for Taiwanese, was promulgated by the MOE in October 14, 2006. Han characters with MOE's Tai-lo Pheng-im are adopted by the CTLT for tests.

The TPT reading tests are divided into two parts: a) vocabulary and grammar and b) reading comprehensions. Readings are arranged in Han-lo style. There are a total of 60 multiple-choice questions. Each question has 4 answer choices, and only one choice is correct. A test-taker will get 3 points if s/he gets a correct answer. In contrast, s/he will be deducted 1 point if s/he gets a wrong answer. No point will be added or deducted if s/he does not answer the questions. The test items of vocabulary and grammar comprise A1, A2, B1, B2, C1, and C2 levels. As for the reading comprehension questions, they comprise B1, B2, C1, and C2

levels.

The listening tests are divided into two parts too: a) conversations and b) talks and lectures. There are a total of 40 multiple-choice questions. The calculation of the score is the same as for the reading test. Conversations refer to dialogues between two or more people. They comprise A1, A2, B1, B2, C1, and C2 levels. Talks and lectures refer to individual talks and lectures on some topics (such as weather reports, story telling, class lectures, and professional speeches). They are expected to be at intermediate and higher levels, so they comprise B1, B2, C1, and C2 levels.

Instead of writing tests, dictation tests in Tai-lo Pheng-im are arranged specifically for the Taiwanese language. The major reasons are: 1) written Taiwanese is currently neither widespread nor standardized. There are several ways to write in Taiwanese. To avoid disagreements over the writing criteria, it is better to exclude composition. 2) Tai-lo Pheng-im is taught in Taiwanese classes. Besides, Tai-lo Pheng-im is a fundamental tool to writing in Taiwanese.

There are 40 Taiwanese words in the dictation tests. The words consists of all consonants, simple vowels, and tones in Taiwanese. They are all tape recorded in advance. Test-takers are asked to write down what they hear in Tai-lo Pheng-im. Each word is repeated three times, and then 3 seconds are left for writing. Scores are calculated in accordance with the percentage of correct phonemes and tonemes the test-taker perceives. For example, assuming that there are a total of 350 phonemes and tonemes in the word list. A test-taker will get 68.6 points ($=80 \times (300/350)$) if s/he get 300 correct phonemes and tonemes.

The speaking tests consist of 3 parts: a) Using picture prompts for storytelling, b) oral readings, and c) oral expressions. There are 2 types of storytelling. The first type is a single picture with some concrete objects. The test-taker is told to describe the contents in simple words. This type of storytelling is classified as an A1 level question. The second type of storytelling comprises 4 pictures in series. Test-takers have to give a short talk in simple ways on the pictures. This type is classified as an A2 level question. For each question in both

types, test-takers have 30 seconds to prepare and 1 minute to record.

For oral readings, there are 2 prepared paragraphs (B1 and B2 levels). Each paragraph is written in Han-lo style and about 300 words long. Test-takers are told to read the paragraphs aloud as fluently as they can. They have 30 seconds to prepare and 2 minutes to record.

As for oral expressions, there are 2 prepared questions (C1 and C2 levels) requesting the test-taker's opinions and ideas on some issues. Test-takers have to express their opinions and ideas fluently. They have 1 minute to prepare and 2 minutes to record.

4. Statistic results of preliminary tests

There were several preliminary tests, and the latest test was held on November 29, 2008. In the latest test, a total of 462 volunteers registered and their actual attendance was 82.3% (380). They were divided into 3 groups and were tested in Tainan, Taichung, and Taipei, respectively. The backgrounds and scores of the subjects are as follows:

Table 2.

Background of preliminary test volunteers (n = 462)

Occupation	Number	Percentage
Elementary school teachers	125	27.1%
Teaching assistants in elementary schools	151	32.7%
General publics	69	14.9%
Professional in Taiwanese (e.g. writers)	14	3.0%
Students	75	16.2%
Others	28	6.1%
Total	462	100%

Table 3. Mean scores of the preliminary test-takers (n = 380)

Sections	N*	Means	Max.	s.d.
Reading	378	109.25	180	24.81
(a) Vocabulary and grammar	378	69.50	108	16.55
(b) Reading comprehension	378	39.75	72	12.71
Listening	380	79.29	120	17.89
(a) Conversations	380	52.29	72	11.15
(b) Talks and lectures	380	27.01	48	9.45
Dictation	380	54.30	80	23.27
Speaking	365	82.89	120	18.80
(a) Picture telling	365	30.25	40	6.45
(b) Oral readings	365	25.00	40	7.59
(c) Oral expressions	365	27.64	40	7.43
Total	363	327.99	500	61.30

* N is the actual number completed for each section.

As mentioned earlier, the TPT is a criterion-reference test (CRT). Therefore, CRT statistical approach should be employed for the TPT. However, the subjects for the preliminary tests were not from the same group. In addition, there was no course given to the subjects to distinguish pre-test and post-test scores. Consequently, norm-referenced (NRT) statistic approach was adopted for the researchers' reference. Readers should be cautious of the statistical differences between CRT and NRT.⁶

The reliability of test was calculated according to Cronbach's α . The results show that internal consistency of the reading section is 0.827, and listening section is 0.771. The overall Cronbach's α for both reading and listening tests is 0.873. These figures show that the items in the TPT listening and reading tests were highly reliable and consistent.

Item facility (IF), also called item difficulty, was calculated according the following formulas:

$$IF = N_{\text{correct}} / N_{\text{total}}$$

N_{correct} = number of subjects who answered correctly

N_{total} = number of subjects taking the test

A P value was further calculated and adapted as an item facility index based on the following formulas:

⁶ For the differences, readers may refer to Brown & Hudson (2002).

$$P = (IF_{upper} + IF_{lower})/2$$

IF_{upper} = item facility for the upper group (1/3 of the total) on the whole test

IF_{lower} = item facility for the lower group (1/3 of the total) on the whole test

In addition to P value, Δ value was also calculated in accordance with Fan's item analysis table for readers' reference (Fan, 1952).

As for the item discrimination (ID) index, it was calculated by subtracting the IF for the lower group from the IF for the upper group as follows (Brown & Hudson, 2002, pp.116-118):

$$ID = IF_{upper} - IF_{lower}$$

Item facility and discrimination statistics are listed in TABLE 4, TABLE 5 and TABLE 6. Values in TABLE 4 were calculated for both reading and listening tests for readers' overall view of the TPT. For the p value, the smaller the value the more difficult it is. As for the Δ value, the higher the value the more difficult it is. So, it is expected that the p value should decrease and the Δ value should increase, from A1 to C2. However, the results showed some exceptions. For example, in TABLE 4, the p value of B2 is unexpectedly higher than B1. Nevertheless, if we reduce the levels to only three, the p value will decrease from A to C. The results reveal some possibilities: 1) the test items in the preliminary test were good enough to distinguish three rather than six levels, or 2) it was the results of statistical errors since the average number of test items in each level was only sixteen. The statistical results might be improved if the number of test items was increased. Further investigation is needed to find the answer.

Item discrimination (ID) index and difference index⁷ (DI) are usually calculated for norm-referenced tests and criterion-referenced tests, respectively (Brown et al., 2002). In TABLE 4, the ID values range from 0.16 to 0.30, with an average of 0.22. The reason for not having a high ID could be that the Taiwanese language ability of the preliminary test volunteers was rather even. According to my observation, only persons who possess a sufficient level of

⁷ The difference index is calculated by subtracting the proportion of the non-mastery group answering the item correctly from the proportion of the mastery group answering the item correctly (Brown et al., 2002, p.120). Difference index of 0.20 and above is considered acceptable (Brown et al., 2002, p.122).

ability in Taiwanese were willing to take the test. Although the ID values are not considered high, it does not mean that the test items were invalid. If we regard the upper group (1/3 of the total) as mastery-level speakers, and the lower group (1/3 of the total) as non-mastery-level speakers, the DI value could be the same as ID value. If so, most ID values of TABLE 4 are higher than 0.20, which are considered acceptable.

Table 4. Item facility and discrimination statistics
on both reading and listening tests

Levels	Item facility index		Item discrimination index
	P	Δ	ID
C2	0.48	13.21	0.26
C1	0.63	11.63	0.30
B2	0.80	9.63	0.26
B1	0.74	10.49	0.22
A2	0.92	7.35	0.13
A1	0.89	8.06	0.16

Table 5. Item facility and discrimination statistics on reading tests

Levels	Item facility index		Item discrimination index
	P	Δ	ID
C2	0.47	13.32	0.30
C1	0.64	11.59	0.30
B2	0.77	10.09	0.33
B1	0.67	11.19	0.25
A2	0.91	7.75	0.17
A1	0.88	8.26	0.17

Table 6. Item facility and discrimination statistics on listening tests

Levels	Item facility index		Item discrimination index
	P	Δ	ID
C2	0.50	13.03	0.19
C1	0.63	11.71	0.30
B2	0.86	8.73	0.15
B1	0.84	9.04	0.17
A2	0.93	6.88	0.08
A1	0.90	7.85	0.15

In the design of TPT, test-takers take the test once and they are assigned a language level according to their scores. Therefore, it is necessary to investigate the relationship between scores and language levels. The relationship was calculated by our team by running ordinal logistic regression based on the testing results of sampled test-takers (Chang, Tu, and Chang, 2009).

For this analysis, data were taken from 160 sampled test-takers, who were included in the 462 volunteers and participated in the preliminary test. Prior to the preliminary test, the sampled test-takers were interviewed by five researchers of the research team and assigned language levels in accordance with the CEF criteria. It would be much better if all 462 volunteers were interviewed. However, due to the project's time limitation, only 160 were interviewed. The sampled test-takers play an important role in running the ordinal logistic regression. We need to use the scores of test-takers to double check whether or not the test items are valid for differentiating language levels.

The number of these sampled test-takers who are at different proficiency levels is listed in TABLE 7. Their mean scores are listed in TABLE 8 and shown in FIGURE 1. FIGURE 1 shows our expectation that the scores in all sections significantly increase from A1 to C2. Among the four sections, the listening has the feature of low slope. After rechecking the testing procedure and test-takers' background, we found the potential factors as follows: In the listening test, the test-takers were told to listen to the recorded passage and then mark their answers on a computer formatted answer sheet with a B2 pencil. No extra writing paper of the answer choices was provided. It was not an easy job for the elderly test-takers to complete the listening test in such conditions.⁸ We further checked their scores and found that the older the test-taker, the more likely for her/him to have lower listening scores than reading scores. In contrast, the collegian test-takers who are familiar with testing skills are more likely to have better listening scores. To solve this problem, an extra sheet of writing paper for working out answer

⁸ Among the test-takers, their age ranges from 77 to 10, and the average age is 42.

choices was provided and the test-takers were allowed to make any notes on it in the first official test in November 2009.

Table 7. Sampled test-takers of different levels

levels	C2	C1	B2	B1	A2	A1	Total
N	8	23	38	39	20	32	160*

* The actual number attended and completed the test is 130

Table 8. Mean scores of sampled test-takers

Levels		Reading	Listening	Dictation	Speaking	Total
C2	mean	148.60	102.40	76.18	101.20	428.38
	s.d.	9.53	3.58	2.70	6.53	13.56
C1	mean	136.81	92.76	71.00	97.95	398.53
	s.d.	8.42	13.62	9.10	9.92	18.89
B2	mean	121.18	79.89	67.39	91.68	360.14
	s.d.	7.96	14.81	11.60	12.80	18.92
B1	mean	105.56	76.22	51.17	80.82	313.77
	s.d.	14.03	21.67	21.93	15.44	26.84
A2	mean	84.37	72.70	41.34	67.11	265.51
	s.d.	20.57	18.90	21.41	17.65	34.38
A1	mean	71.29	64	23.08	41.43	199.79
	s.d.	21.87	12.82	16.81	13.88	43.57
Total	mean	107.05	78.56	53.53	79.30	318.45
	s.d.	27.04	19.30	23.15	22.29	69.58

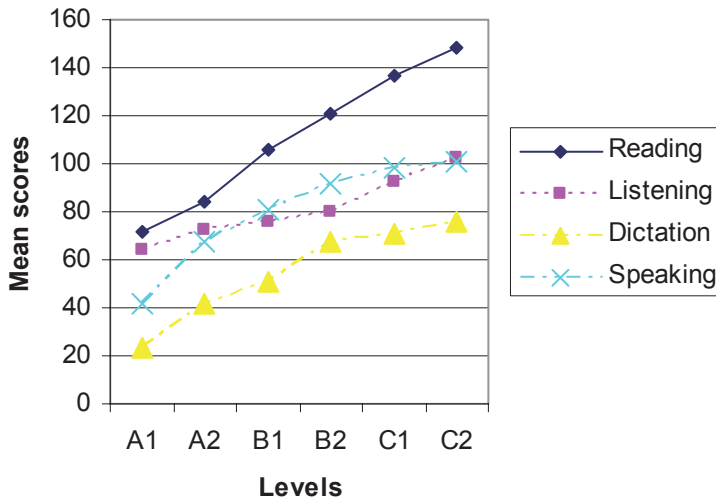


Figure 1. Mean scores of sampled test-takers.

The probabilities of the sampled test-takers' language levels were calculated according to their scores by using ordinal logistic regression as shown in FIGURE 2. For example, the node of the two left-most curves is 220. If a sampled test-taker receives a grade less than 220, s/he is more likely to be regarded as A1 level. In contrast, if s/he receives a score between 220 and 290 (the node between the second and third curve), s/he is more likely to be A2 level.

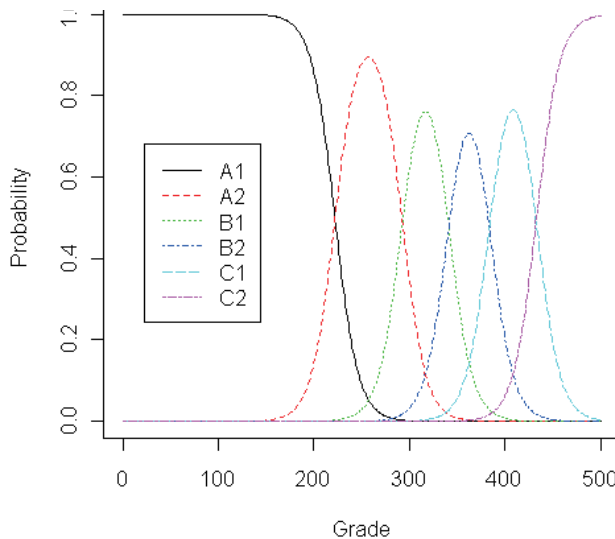


Figure 2. Probability of sampled test-takers' language levels.

According to the results of ordinal logistic regression, the nodes made by the curves are at 220, 290, 340, 380, and 430, respectively. They are treated as the boundary scores between different language levels, as shown in TABLE 9.

As for the minimum scores of A1, the range (70=290-220) between A2 and A1 is treated as the range between A1 and zero levels. Therefore, the minimum scores of A1 is 150 (=220-70).

Table 9.
 Contrasts between scores and language levels

Levels	Scores
C2 Mastery	$430 < \text{scores} \leq 500$
C1 Effective Operational Proficiency	$380 < \text{scores} \leq 430$
B2 Vantage	$340 < \text{scores} \leq 380$
B1 Threshold	$290 < \text{scores} \leq 340$
A2 <u>Waystage</u>	$220 < \text{scores} \leq 290$
A1 Breakthrough	$150 < \text{scores} \leq 220$

The accuracy of rating scale of language levels on scores was calculated and obtained as 0.68 ($= (9+18+21+22+15+3)/130$), as shown in TABLE 10. The number of sampled test-takers of different language levels assigned in advance by the researchers was listed in the column as “total.” The row “total” is the forecasted number of different language levels according to the scores after test of sampled test-takers. For example, in the column of A1, 4 subjects who were evaluated as A2 level before test, were forecasted to be in the A1 level according

to their scores after test. In the same column, 9 subjects were graded as A1 both before and after test. The results reveal that TPT's judgment on subjects' language levels reach a 0.68 accuracy, which is much higher than the probability 0.17 (=1/6) by guessing.

Table 10. Cross table of assigned and forecasted language levels

assigned \	Forecasted levels by actual scores						total
	A1	A2	B1	B2	C1	C2	
C2	0	0	0	0	2	3	5
C1	0	0	0	6	15	0	21
B2	0	0	4	22	2	0	28
B1	0	8	21	3	0	0	32
A2	4	18	8	0	0	0	30
A1	9	5	0	0	0	0	14
total	13	31	33	31	19	3	130

In the TPT, a test-taker will be assigned a language level only if s/he completed all four sections of the tests. When the same scoring scheme was applied to all the subjects who participated in the preliminary test in November 2008, the results reveal that 46.32% of the subjects received level B2 or higher (see TABLE 11). This result meets the expectation as B2 is the recommended level by CTLT as the minimum Taiwanese ability requirement for teaching Taiwanese in elementary schools.

Table 11. Levels obtained by all subjects

Levels	N	%	Accumulated %
C2	6	1.58	1.58
C1	62	16.32	17.90
B2	108	28.42	46.32
B1	98	25.79	72.11
A2	73	19.21	91.32
A1	26	6.84	98.16
Less than A1	7	1.84	100
total	380	100	

5. Conclusion

The Taiwanese Proficiency Test (TPT) is a newly developed language testing scheme for educational purpose. In the test, test-takers are assigned a language level according to the scores they receive. The relationship between language proficiency levels and scores is calculated based on the statistical results of 160 sampled test-takers by using ordinal logistic regression. Its accuracy for predicting test-takers' language levels reaches as high as 0.68. The accuracy may not be perfect. However, it probably is the best that can be obtained under time and resource constraints. The first official TPT was held by the CTLT with a total of 793 test-takers on November 14, 2009. In addition, more trial tests are proposed to be conducted later. Accuracy is expected to improve gradually with the employment of official and trial tests in the near future. Hopefully, the TPT will benefit the teaching of Taiwanese language, and will further empower the revival of Taiwanese.

Acknowledgments

This paper is a result of research project (period: 2008.4.17-2009.4.16) sponsored by the National Languages Committee of Taiwan's Ministry of Education. As the coordinator of the project, I am indebted to all participants who were involved in the research project. The TPT would not be possible without their contributions.

Correspondence

Any correspondence should be directed to Wi-vun Taiffalo Chiung (uibuntw@gmail.com).

References

- Brown, James D., and Thom Hudson. 2002. *Criterion-referenced language testing*. Cambridge: Cambridge University Press.
- Chan, Hui-Chen 1994. *Language shift in Taiwan: social and political determinants*. Ph.D. dissertation: Georgetown University.
- Chang, L.C.; Y.H. Tu; and S.M. Chang. 2009. 統計方法於試題研發之應用 [Statistic Approaches for the Taiwanese Proficiency Test]. Proceedings of the Second Conference on Taiwanese Proficiency Test (March). Tainan, Taiwan: National Cheng Kung University.
- Cheng, Robert L. 1996. 民主化政治目標kap語言政策 [Democracy and language policy]. In C.H. Si (pp. 21-50).
- Chiung, Wi-vun T. 2001. Language attitudes toward written Taiwanese. *Journal of Multilingual and Multicultural Development*, 22(6), 502-523.
- Chiung, Wi-vun T. 2006. 全民台檢第一擺試考結果統計分析 [Statistical analysis of the first GTPT preliminary test]. Proceedings of the First Conference on Taiwanese Proficiency Test (December). Tainan, Taiwan: Tainan Theological College and Seminary.
- Fan, C.T. 1952. *Item analysis table*. Princeton: Educational Testing Service.
- Hsiau, A-chin. 1997. Language ideology in Taiwan: the KMT's language policy, the Tai-yu language movement, and ethnic politics. *Journal of Multilingual and Multicultural Development* 18 (4), 302-315.
- Huang, Shuan F. 1993. 語言社會與族群意識 [Language, Society, and Ethnic Identity]. Taipei: Crane.
- Kerr, George H. 1992. 被出賣的台灣 [Formosa Betrayed]. Taipei: Chian-ui Press.
- McNamara, Tim. 2000. *Language testing*. Oxford: Oxford University Press.

- Ong, Iok-tek. 1993. 台灣苦悶的歷史 [*Taiwan: a depressed history*]. Taipei: Independence Press.
- Peng, Ming M., and Yuzin C. Ng. 1995. 台灣在國際法上的地位 [*The legal status of Taiwan*]. Taipei: Taiwan Interminds Publishing Inc.
- Si, Cheng-hong. (ed.) 1996. 語言政治與政策 [*Linguistic politics and policy*]. Taipei: Chian-ui.
- Su, Beng. 1980. 台灣四百年史 [*Taiwan's 400 year history*]. San Jose: Paradise Culture Associates.
- Teng, Hong-tin; Le-kun Tan; Jin-sek Goo; Ju-khai Lim; Sin-han Ho; and Wi-vun Chiung. 2006. 全民台語能力檢定測驗規劃ê現況 [A survey on current General Taiwanese Proficiency Test]. Proceedings of the First Conference on Taiwanese Proficiency Test (December, pp. 53-63). Tainan, Taiwan: Tainan Theological College and Seminary.
- Tiun, Juhong. 1996. 台灣現行語言政策動機ê分析 [An analysis on the Taiwan's current language policy]. In Cheng-hong Si (pp. 85-106).
- Young, Russell. 1988. Language maintenance and language shift in Taiwan. *Journal of Multilingual and Multicultural Development* 9 (4), 323-338.