

國小2-3年級團體智力測驗之編製 及其信效度研究

蔡明富

國立高雄師範大學特殊教育學系

摘 要

本研究旨在編製三式國小2-3年級團體智力測驗（Group Intelligence Test for Children in Grade 2-3, 簡稱GITC-G23），以供資優學生鑑定的初步篩選工具。本測驗主要根據流體-晶體智力理論、評量內容兼具語文與非語文、節省施測時間及多套複本測驗等四大特色發展本測驗。本測驗包含語文反義、數字系列與圖形類比三個分測驗，全測驗共有59題，作答採五選一方式，施測時間約40分鐘（含作答說明）。研究結果顯示：在信度方面，分測驗內部一致性係數Cronbach α 係數介於 .712至 .835之間，全測驗為 .882。分測驗間隔四週重測信度係數介於 .724至 .885之間，全測驗為 .893；三式測驗複本信度係數介於 .725至 .794之間。在效度方面，除具內容效度外，本測驗分別與學業成績、「國民小學語文智力測驗」及「托尼非語文智力測驗」有不錯的效標關聯效度；本測驗亦能有效區別資優生與普通生；驗證性因素分析結果得到晶體智力與流體智力兩個因素，顯示本測驗與原先建構因素相符。最後，以2,135名國小2-3年級學生為常模樣本，建立甲式測驗百分等級及常態化的T分數常模。並根據研究結果提出建議，以供未來研究參考。

關鍵詞：資優學生、團體智力測驗、信度、效度

蔡明富

高雄市苓雅區和平一路116號（國立高雄師範大學特教系）

(07)717-2930#2325

mingfu@nknuc.nknu.edu.tw

壹、緒論

一、研究動機

由於資優學生智力發展快速，讓他們認知發展有別於學校的普通班同儕（Clark, 2012; Preuss & Dubow, 2004）。故相較於普通班同儕，資優學生具有優於普通學生的學習能力，為因應其特殊教育需求，如何適時發掘資優學生甚為重要。目前針對資優學生鑑定，鑑定工具的發展仍有待突破。根據教育部（2008）公布《資優教育白皮書》，在「鑑定與安置」章節明白指出：「國內目前資優鑑定工具缺乏，部分縣市之鑑定過程不夠嚴謹，以致測驗內容外洩，一方面嚴重影響資優鑑定的效度，另一方面也突顯資優鑑定工具的不足。」因此，研發資優學生鑑定評量工具乃為未來重要發展策略之一。目前國內資優學生鑑定過程中，智力測驗扮演相當重要角色。智力測驗施測方式可分成團體與個別測驗，團體測驗在格式及題目的安排與個別測驗不同，團體測驗有助於團體施測用的工具，若與個別測驗相較，它的優點包括：可同時多人一起施測、施測容易、簡化主試者的角色、計分較客觀、及易建立常模（Kaplan & Saccuzzo, 2013; Murphy & Davidshofer, 2004）。由於國內各縣市在國民小學資優學生鑑定時，均需要先進行初步篩選機制，故團體智力測驗能達到大量施測且初步篩選資優學生之用，以利後續個別智力測驗之施測，可有效節省施測時間及人力。為解決資優學生鑑定之初步篩選工作，團體智力測驗是一項重要且不可或缺的鑑定工具。

目前編製智力測驗多數仍以因素分析取向為主，依據智力因素構念作為測驗的理論架構（Carroll, 1993; Demetriou & Efklides, 1994; Kamphaus, Petoskey, & Morgan, 1997）。是故，本測驗乃根據傳統心理測驗

學者所提出若干重要智力因素進行編製。在傳統智力理論中，Cattell與Horn認為個體認知能力包括流體智力（fluid Intelligence, Gf）與晶體智力（crystallized Intelligence, Gc），提出流體-晶體智力理論（Gf-Gc theory）最受重視（Horn & Blankson, 2005）。過去許多頗具盛名智力測驗工具，編製基礎以參酌Cattell-Horn-Carroll（CHC）理論為主，惟編製者會依其測驗性質採取不同因素建構測驗架構，其中最常被採納因素包括晶體智力與流體智力（Flanagan & McGrew, 1997; Stankov, 1991）。目前對智力定義乃相當多元，考量團體智力測驗施測時間有限，本測驗編製目的不在測量廣泛之「智力」，乃參酌CHC理論中兩項重要因素：晶體智力與流體智力作為編製本測驗主要架構依據。如此可使測量的構念更明確，且易於實施。另根據過去編製測驗內容除語文測驗外，亦應包括非語文測驗，以避免地區、文字材料和教育文化背景等條件對學生智力評估之影響（Bracken & Naglieri, 2003; Hussain, Jamil, Siraji, & Maroof, 2012）。故本測驗編製內容會兼顧語文測驗及非語文測驗。

古典測驗理論（classic test theory, CTT）乃過去編製測驗較常採用，雖然其理論易被大家接受，惟許多學者（de Ayala, 2009; Hambleton & Swaminathan, 1985）指出古典測驗理論會出現以下缺點，例如：(1)題目參數（難度及鑑別度）會隨著樣本特質而改變；(2)無法事先估計不同能力受試者在測驗的可能表現；(3)不易編製平行測驗；(4)無法進行測驗分數的等化；(5)未能提供不同能力受試者可能答對每個試題機率之函數。試題反應理論（item response theory, IRT）的出現可改善古典測驗理論的缺失，因為IRT採取題目訊息（item information）、測驗訊息（test information）及能力估計值標準誤（standard error）作為評鑑測驗選題的指標，這三類訊

息屬於能力的函數而非常數。本測驗的編製過程將納入試題反應理論的觀點，經由水平等化取得所有試題的鑑別度、難度及猜測度三參數，再依據目標訊息函數選取適合的題目來組成所要的測驗，故本測驗的編製過程會納入試題反應理論的觀點，期使本測驗的編製更具嚴謹。

二、研究目的

由於國內資優學生鑑定工具有限，如能編製團體智力測驗可有效初步篩選資優學生。為使教育現場能夠更有效率評量學生認知能力，研究者考量智能評量重要元素應以晶體智力及流體智力為主，測驗內容應納入語文與非語文測驗，另為避免題目外流，且增加測驗使用年限，有必要編製多套測驗。另為達到測驗編製歷程更精確，在測驗發展過程將採用試題反應理論組合測驗。因此，本測驗發展過程採用三參數試題反應理論來進行題目參數連結，並根據題目參數組成測驗訊息曲線非常接近的三式複本測驗。故本研究目的乃試圖編製三式盡可能接近平行測驗之要求的「國小2-3年級團體智力測驗」(group intelligence test for children in grades 2-3, GITC-G23)，以供篩選國小資優學生。

貳、文獻探討

一、資優學生鑑定與團體智力測驗

依據教育部2012年6月修正發布《特殊教育學生調整入學年齡及修業年限實施辦法》，第2條指出調整修業年限資優學生包括：降低入學年齡學生（指提早入國民小學就讀之年齡），以及縮短修業年限學生，指縮短專長學科（學習領域）學習年限或各教育階段修業年限。另第3條指出：前項資賦優異兒童之鑑定，應符合下列規定：(1)智能評量之結果，在平均數正二個標準差以上或

百分等級九十七以上。(2)社會適應行為之評量結果與適齡兒童相當。顯示調整修業年限資優學生需要進行智能評量。另依據教育部2013年9月修正發布《身心障礙及資賦優異學生鑑定辦法》，第2條指出：「資賦優異學生之鑑定，應以標準化評量工具，採多元及多階段評量，除一般智能及學術性向資賦優異學生之鑑定外，其他各類資賦優異學生之鑑定，均不得施以學科（領域）成就測驗。」另在第15條指出：「一般智能資賦優異，指在記憶、理解、分析、綜合、推理及評鑑等方面，較同年齡者具有卓越潛能或傑出表現者。」前項所定一般智能資賦優異，其鑑定基準依下列各款規定：(1)個別智力測驗評量結果在平均數正二個標準差或百分等級九十七以上。(2)經專家學者、指導教師或家長觀察推薦，並檢附學習特質與表現卓越或傑出等之具體資料。故一般智能資優學生的鑑定需要進行個別智力測驗。從以上條文發現，目前國內縮短修業年限資優學生及一般智能資優學生的鑑定，均需要智能評量，突顯出智能評量在資優學生鑑定過程，扮演重要角色。

Cooper (1999) 認為團體測驗與個別測驗相比較，團體測驗的優點為施測容易、省時及較個別測驗更容易建立常模，顯示團體智力測驗有其重要性。教育部已於2007年1月18日公告之「國民教育階段資優學生鑑定安置流程」中，指出資賦優異學生之鑑定，應以標準化評量工具，採多元及多階段之評量方式，其評量之實施應依觀察、推薦、初審、初選、複選及綜合研判之程序辦理。更明白指出國民小學一般智能優異學生在初選時需要團體智力測驗，接著複選階段再進行個別智力測驗。目前國內國小一般智能資優班的鑑定時程以國小二年級下學期為主，顯示發展國小二年級資優學生初步篩選工具有其迫切性。因此，本研究主要目的在編製適合篩選國小2-3年級資優生之團體智力測驗。

二、智力理論的探討

Carroll (1993) 指出目前心理學家對智力定義相當多元，有的學者認為智力是學習新事物的能力或利用經驗的能力；有的學者主張智力是抽象思考與判斷的能力；也有些學者認為智力是適應新環境的能力；另有的學者以學校教育的立場說明智力是學習學校內各種工作的能力。目前關於智力的構念，可分成傳統智力理論與新智力理論，在傳統智力理論中，許多學者提出不同觀點，如Thorndike的「多因論」強調：「智力是由若干個分割的因素構成，每個因素即為一個小的能量單元」（引自Cooper, 1999; 引自Kline, 1991）。Spearman「二因論」則根據多年測驗和統計結果，證明人類智力的高低為兩個因素所決定：一為各種心智活動所共同具有的「一般因子」(G)，二為學習各種專門業務所依據的「特殊因子」(S)。Thurstone「群因論」則認為智力是由幾種「基本能力」所組成，他根據多年測驗的結果將基本能力歸納為七種：數學、語文、用字流利、記憶、推理、空間以及知覺速度等能力 (Cooper, 1999; Kline, 1991)。另從新智力理論觀點，Swanson (1985) 認為智力成分可分為兩大類：一為智力的認知成分；二為智力的執行成分。其中智力的認知成分著重於人的認知限度（例如：知覺儲存、短期記憶、及長期記憶等）；智力執行成分重視人類處理訊息或認知歷程的運作及使用策略的功能等。新智力理論觀點的學者，例如Sternberg (1985) 主張智力三元論 (Triarchic Theory of Intelligence)，其將智力分成「內在的成份」(componential)、「外在的環境」(contextual) 以及交互的經驗 (experiential)。Naglieri 與 Das (1997) 則提出訊息處理理論，將智力視為心智技能，是處理訊息及解決問題的能力。然而依據新智力理論編製評量工具仍易產生困難，

Sternberg (1992) 認為目前依據新智力理論所發展出來的測驗工具，僅能補充傳統測驗之不足，但仍無法取代之。因此，依據傳統智力理論所編製之測驗，仍有其價值。是故，本測驗乃根據傳統心理測驗學者所提出若干重要智力因素進行編製。

在傳統智力理論中，首先以Catell與Horn提出流體-晶體智力理論最受重視，晶體智力係指個體在所處文化環境下的學習知識，並用以解決相關問題，故晶體智力與個體接受教育與語文能力有很大關係；流體智力係指具有解決問題的能力，且解決的這些問題是未接受教育及文化層面影響 (Horn & Blankson, 2005)。之後Cattell-Horn-Carroll結合流體-晶體智力理論及Carroll (2005) 主張三階層認知理論模式 (the three-stratum theory of cognitive abilities)，提出CHC智力理論，其將評量智力內涵分為10項：晶體智力 (crystallized intelligence, Gc)、流體智力 (fluid intelligence, Gf)、計量知識 (quantitative knowledge, Gq)、讀寫能力 (reading and writing ability, Grw)、短期記憶 (short-term memory, Gsm)、視覺處理 (visual processing, Gv)、聽覺處理 (auditory processing, Ga)、長期記憶儲存與提取 (long-term storage and retrieval, Glr)、處理速度 (processing speed, Gs)、決定/反應時間或速度 (decision/reaction time or speed, Gt) (Wilhoit & McCallum, 2003)。其中晶體智力泛指對文化資訊廣度與深度的認識，而這些資訊是個體需要與運用的重要知識。Gc包括：語言發展 (language development)、語彙知識 (lexical knowledge)、聽力 (listening ability)、常識 (general information)、文化資訊 (information about culture)、科普資訊 (general science information)、溝通能力 (communication ability)、口說與流

暢能力 (oral production and fluency) 等等。流體智力指因應新事物的問題解決能力。Gf包含：一般排序推理 (general sequential reasoning)、歸納 (induction)、皮亞傑式推理 (piagetian reasoning) 和推理速度 (speed of reasoning)。上述討論發現智力內涵相當多元，要設計一套非常完善智力測驗是不太容易。惟原則上一套有效的智力測驗需廣泛從認知能力的全部範疇中，抽出足以反映智力特性的重要成分，但事實上易受到測驗時間長短的限制而不易達到此目標。

從1995年後已出版國外重要智力測驗內容觀之，均以CHC理論進行編製，其中包括Woodcock-Johnson III Tests of Cognitive Abilities (Woodcock, McGrew, & Mather, 2001)、Wechsler Intelligence Scale for Children (Wechsler, 2003)、Stanford-Binet Intelligence Scales (Roid, 2003)、Kaufman Assessment Battery for Children (Kaufman & Kaufman, 2004)、及Differential Ability Scales (Elliott, 2007)。上述著名之智力測驗工具，編製基礎以參酌CHC理論為主，編製者會依其測驗目的採取不同智力因素以建構測驗，其中最常被採納因素包括晶體智力與流體智力 (Flanagan & McGrew, 1997)。Volker、Lopata與Cook-Cottone (2006) 提及CHC理論可有效評量出資優學生的能力，其中晶體、流體智力是測量高層思考推理的能力，過去一般智能資優學生測驗會納入晶體、流體這兩項能力。綜上所述，在智力定

義相當多元下，且考量有限施測時間，本測驗編製目的不在測量廣泛之「智力」，最後決定本測驗編製主要參酌CHC理論中兩項重要因素：晶體智力與流體智力，如此可使本測驗的構念更精簡、明確，且易於教育現場實施及運用。

三、國內外團體智力測驗工具內容之評述

(一) 國內團體智力測驗之探討

針對國內現有測驗，林幸台 (2002) 曾整理智力測驗在各智能運作方式中採用之題型，歸納出現有智力測驗之題型大致可分為以下四類：(1)記憶：延宕記憶。(2)認知理解：字詞替換、詞義辨別、語詞歸類 (語文歸類)、比較大小 (數學理解)、及數系。(3)推理：語文推理、算術推理。(4)評價：邏輯推理。研究者再自行整理國內現有團體智力測驗評估內容，請見表1所示。從表1得知國內團體智力測驗有的以語文測驗為主 (林幸台, 2002; 黃國彥、鍾思嘉、林珊如、李良哲, 1983)，有的偏重非語文測驗 (朱錦鳳, 2005; 吳武典等人, 1996; 林美珍、林邦傑、黃國彥、李富言, 2002; 徐正穩、路君約, 1995; 黃國彥、鍾思嘉、傅粹馨, 1977; 陳榮華, 1998; 許天威、蕭金土, 1999; 翁儷禎, 1997)，有的兼顧語文及非語文測驗 (吳武典、金瑜、張靖卿, 2010; 陳龍安, 1996; 路君約、盧欽銘, 1991)。

表1

國內團體智力測驗整理表

編製者	測驗名稱	對象	測驗內容
黃國彥、鍾思嘉與傅粹馨(1977)	羅桑二氏非語文智力測驗	小三至大一	測驗內容包括：圖形分類、數系、圖形類推。
黃國彥、鍾思嘉、林珊如與李良哲(1983)	羅桑二氏語文智力測驗	小三至高三	本測驗有第一、第二兩式複本，每式均有五個分測驗：字彙、語句完成、算術推理、歸類、語文類推等測驗。
路君約與盧欽銘(1991)	國民中學智力測驗第三種	國中學生	1.語文部份：包括語文類推、語文歸納測驗 2.數學部份：包括算術計算、算術推理測驗
俞筱鈞修訂(1993)	瑞文氏圖形推理測驗系列一彩色(CPM)	6½歲至9½歲	測驗內容包括三種類型：歸納、類推、填充。
俞筱鈞修訂(1993)	瑞文氏圖形推理測驗系列一標準(SPM)	9½歲至12½歲	測驗內容包括三種類型：歸納、類推、填充。
徐正穩與路君約(1995)	圖形式智力測驗	小學三年級至成年人	測驗內容包括：歸納測驗、類推測驗、填充測驗。
吳武典等人(1996)	托尼非語文智力測驗(TONI)_幼兒版	4歲至7½歲	測驗包括：簡單配合；相似性(相等、相加、相減、改變、漸進)；分類；交叉；漸進。
吳武典等人(1996)	托尼非語文智力測驗(TONI)_普及版	7½歲至16½歲	測驗包括：簡單配合；相似性(相等、相加、相減、改變、漸進)；分類；交叉；漸進。
陳龍安(1996)	智能結構學習能力測驗	幼稚園大班至小學三年級	認圖單、認圖類、認語單、評圖單、聚圖單、認語關、認語系、記符單、認符系、記圖單
翁儷禎(1997)	推理思考測驗	高中生或成人	測驗內容包括：圖形類比、圖形序列、語文類比、反義測驗、邏輯推理、數系測驗。
陳榮華(1998)	陳氏非語文能力測驗	12歲以上	測試的面向包括記憶、知覺速度、抽象推理、空間關係等，題型包括：知覺速度測驗、點數推算測驗、空間辨認測驗
許天威與蕭金土(1999)	綜合性非語文智力測驗	6歲至14歲	1.類比推理能力：物品圖畫類比測驗、幾何圖形類比測驗 2.歸類區分能力：物品圖畫歸類測驗、幾何圖形歸類測驗 3.排序推理能力：物品圖畫排序測驗、幾何圖形排序測驗
林幸台(2002)	國民小學團體語文智力測驗	7至12歲	1.甲、乙式：歸類測驗、算術推理、延宕記憶、語文類推、數字系列、邏輯推理。 2.丙式：歸類測驗、算術能力、延宕記憶、比大小(數字)、比大小(圖形)、語文類推。

(續下頁)

表1 (續)

編製者	測驗名稱	對象	測驗內容
林美珍等人 (2002)	國民小學團體 非語文智力測 驗	7至10歲	測驗內容包括：整合辨同、歸納、類推。
朱錦鳳(2005)	圖形思考智能 測驗	適用於大專至 成人	測驗內容包括：點線描繪、形狀組合、方格分解。
吳武典、金 瑜與張靖卿 (2010)	多向度團體智 力測驗兒童版	小二至小六	1.語文部分：類同、理解、常識、詞彙、數學。 2.非語文部分：辨異、排列、空間、拼配、譯碼。

(二) 國外團體智力測驗之探討

研究者整理國外現有團體智力測驗評估內容，請見表2。從表2顯示國外團體智力測驗多數以非語文測驗為主（Brown, Sherbenou, & Johnsen, 1997; Hammill, 表2

Pearson, & Wiederholt, 1997; Kamphaus, 1993; McCallum & Bracken, 1998; Naglieri, 2003; Naglieri & Bardos, 1997），有的兼採語文及非語文測驗（Lohman & Hagen, 2001; Otis & Lennon, 1996）。

表2
國外團體智力測驗整理表

編製者	測驗名稱	對象	測驗內容
Otis & Lennon(1996)	Otis-Lennon school ability test	幼稚園至高中	1.語文部分：語文理解、語文推理、計量推理。 2.非語文部分：圖像推理、圖形推理
Brown, Sherbenou, & Johnsen(1997)	Test of Nonverbal Intelligence	6至89歲	測驗內容包括：類化與分類、區分、類比推理、系列化、歸納、推論、細節確認。
Naglieri & Bardos(1997)	The General Ability Measure for Adults	18歲以上	主要是了解受測者的推理與邏輯解題能力，測驗內容包含四個類別：圖形配對、圖形類比、圖形排序、圖形建構。
Hammill, Pearson, & Wiederholt (1997)	Compre- hensive Test of Nonverbal Intelligence	6至89歲	測驗內容包括：類比、歸納、排序。
McCallum & Bracker(1998)	The Universal Nonverbal Intelligence Test	5至17歲	測驗內容包括： 1.符號測驗：記憶部份（符號記憶、物體記憶）、推理部份（類比推理）； 2.非符號測驗：記憶部份（空間記憶）； 3.推理部份（立方體設計、迷宮）

(續下頁)

表2 (續)

編製者	測驗名稱	對象	測驗內容
Lohman & Hagen (2001)	Cognitive Abilities Test	幼稚園至高中	1.幼稚園至小三： 測驗主要分為三個部份：語文（口語字彙、字彙推理）；計量(計量概念、數字關係)；非語文（數字、圖形分類）。 2.小四至高中： 測驗主要分為三個部份：語文（句子完成、字彙分類、字彙類推）；計量(數量關係、數序及方程式建立)；非語文（圖像分類、圖像類推及圖像分析）。
Naglieri(2003)	Naglieri Nonverbal Ability Test	5至18歲	測驗內容包括：樣式完成、類比推理、系列推論、空間視覺化。

(三) 小結

在表1及表2所探討的國內外團體智力測驗，國內部分團體智力測驗是修訂國外測驗而來，例如：吳武典等人（1996）編訂托尼非語文智力測驗（TONI）主要修訂自Brown、Sherbenou 與 Johnson（1997）的Test of Nonverbal Intelligence，黃國彥、鍾思嘉、林珊如與李良哲（1983）修訂羅桑二氏語文智力測驗，乃屬於Lohman 與 Hagen（2001）編製的Cognitive Abilities Test之前身。從測驗編製內容而言，發現國內外團體智力測驗評估內容除著重晶體智力（例如：語文、數學推理思考能力）外，亦包括流體智力（例如：空間推理能力）等測驗內容。另發現測驗內容除語文測驗外，為避免區域性、文字材料和教育文化背景等條件對學生智力評估之影響（Bracken & Naglieri, 2003; Maller, 2003; McCallum, 2003; Naglieri, 2003），兼採圖形類之非語文測驗，以減少文化背景對施測結果之影響。故智力測驗評估應重視晶體及流體智力外，評量內容亦應包括語文與非語文。本研究目的乃編製以晶體智力（包括語文與數學）及流體智力（圖形）測驗為架構的團體智力測驗，其中語文類、數學類及圖形類分測驗，各分測驗內容選取以上述探討團體智力測驗文獻中常測量的構念為

主，以符合本測驗內容建構具有文獻支持。另從測驗編製理論而言，國內外測驗編製理論係以古典測驗理論為主，鮮採取試題反應理論進行選題與建立複本測驗。本測驗編製過程採取IRT進行測驗選題及建立三式複本測驗。

參、測驗編製歷程

一、測驗之原始架構與設計

根據上述國內外相關文獻之探討，在智力理論中，以Cattell與Horn提出流體-晶體智力理論頗受重視，本測驗編製乃以此一理念為基礎，研究者參酌CHC理論中兩項重要因素：晶體智力與流體智力作為主要編製理論依據。另根據前述文獻討論結果發現國內外智力測驗題型內容編製應同時兼顧語文與非語文測驗。故本研究乃編製以晶體智力（包括語文與數學）及流體智力（圖形）為架構的團體智力測驗，其中語文測驗包括語文類與數學類測驗，分測驗內容選取以過去智力測驗曾測量的構念為主。語文類測驗包括語文反義測驗，數學類測驗包括數字系列測驗，圖形測驗包括圖形類比測驗。本研究編製各分測驗架構請見表3。

表3

GITC-G23各分測驗內容架構表

測驗內容	分測驗名稱
一、語文部分	語文反義測驗
二、數學部分	數字系列測驗（數字運算、數字推算）
三、圖形部分	圖形類比測驗（圖形對照、圖形變化）

二、擬題與確定預試題本

本計畫編製題目可分為初擬題目與學者專家審題兩部份說明：

（一）初擬題目

在確定測驗架構之後，即著手蒐集各種測驗之相關題型，參考其中的項目，編製雙向細目表（two-way specification table）。並邀集南部地區資深國小國語（文）及數理教師各10名參與題目之撰寫，經五次會議討論後，再邀請資深教師進行初步選題，完成國小2-3年級之團體智力測驗題庫。再依各領域命題者之經驗，將題目分成甲、乙、丙三式；每式各有三個分測驗，各式的題數、題型皆相似，且難度相當。

（二）學者專家審查和討論

完成擬題後，接著進行學者專家審查，邀請高雄師範大學、台南大學及屏東教育大學共三名資優領域學者專家進行審查，審查

表4

預試各題型分佈之題數

分測驗	題型	甲式	乙式	丙式
語文反義	語文反義	23	23	23
數字系列	數字運算	15	13	12
	數字推算	8	10	11
圖形類比	圖形對照	12	12	9
	圖形變化	13	13	16

三、進行測驗預試

（一）預試時間評估

預試工作主要是探究主試者施測流程與方式、時間評估及受試者的反應等情形。本測驗之預試時間共分為兩次，第一次預試以高雄市凱旋國小2-3年級各3名學生進行預試時間評估，結果發現語文反義施測時間過

和討論的流程與重點包括：(1)研究者與學者專家分別審查初始題庫題項歸類與編題架構之符合性；(2)檢核審查委員對歸類適切性之意見，逐步依據編題架構，將有共識之題項予以保留，針對意見歧異題項進行逐題討論，經由修改題義、文句、拆拼題項，以達成共識。(3)最後根據學者專家意見予以修改成預試題本。

（三）確定預試題本與共同試題

經過上述審查與討論後形成預試題本各三式，其中每式測驗包含「語文反義」分測驗23題、「數字系列」分測驗23題、「圖形類比」分測驗25題。本研究因採IRT進行題目編製，為有利進行試題水平等化及取得試題的鑑別度、難度和猜測度三參數，故三式之間彼此有共同試題作連結，以利資料分析時進行試題等化。關於甲、乙、丙三式之題型分佈及題數，請見表4。

長、圖形類比施測時間過短，於是將語文反義施測修改為10分鐘、圖形類比修改為15分鐘。第二次預試根據第一次預試結果，以高雄市福康國小2-3年級各兩班的學生為樣本，確定施測時間及施測流程；依據此次預試之受試者反應及時間估計，重新評估及修正各分測驗施測流程及所需時間，請見表5。

表5

GITC-G23預試題本各分測驗名稱、題數及作答時間

分測驗	題數	第一次預試	第二次預試
語文反義	26	11分鐘	10分鐘
數字系列	26	15分鐘	15分鐘
圖形類比	25	14分鐘	15分鐘

(二) 預試樣本

預試樣本以高雄市之國小2-3年級學生為主，共抽取12所學校之學生為預試對象，每校2-3年級各抽取3班。國小2-3年級預試題

本共分為甲、乙、丙三式，其中各式預試樣本數分別為720名、709名、與714名，受試樣本之年級、性別及人數等資料，請見表6。

表6

預試題本施測基本資料(高雄市)

區域	學校名稱	年級	甲式		乙式		丙式	
			男	女	男	女	男	女
楠梓區	加昌國小	二	19	17	19	17	19	17
		三	19	15	18	16	19	15
苓雅區	四維國小	二	17	17	15	14	16	15
		三	16	14	14	14	15	14
鼓山區	九如國小	二	17	13	19	11	17	10
		三	17	15	18	15	19	13
前鎮區	光華國小	二	17	13	17	13	18	13
		三	15	14	14	15	15	14
前金區	前金國小	二	19	16	18	15	15	18
		三	16	12	17	11	8	20
旗津區	旗津國小	二	18	15	16	14	15	16
		三	14	14	13	16	15	14
鳳山區	鳳山國小	二	15	12	16	14	16	13
		三	15	13	15	13	16	13
岡山區	岡山國小	二	18	10	17	13	17	13
		三	15	13	14	14	16	13
鳥松區	鳥松國小	二	15	15	14	15	16	13
		三	14	14	15	14	16	14
林園區	中芸國小	二	11	12	9	14	11	14
		三	17	14	17	14	15	15
橋頭區	仕隆國小	二	17	13	15	12	16	12
		三	12	13	11	12	13	11
大樹區	九曲國小	二	13	18	15	15	15	14
		三	19	13	18	14	20	12
合計			385	335	374	335	378	336

(三) 進行題目分析與選題

本測驗共分成甲、乙、丙三式進行預試，待預試測驗回收後，將每式之分測驗試題建立成一測驗題庫，共有語文反義測驗題

庫、數字系列測驗題庫、及圖形類比測驗題庫。各測驗題庫之試題以項目分析軟體 (iteman) 進行題目分析，計算每一試題的答對率 (P值)、高低分組答對率差距之D值

以及點二系列相關 (r_{nb} 值) 等試題鑑別度。另採試題反應理論進行本測驗之選題，由於IRT的優點是可以進行試題等化，IRT可經由不同測驗間的共同試題作連結，由於共同試題間每個題目難度參數 (b) 有一對分別來自兩個樣本群體的估計值，經由這組數字間的關係可進行線性轉化，使不同試卷的題目參數化為同一量尺單位，如此將可取得不受樣本影響的獨立題目參數，作為測驗最後組題的參考依據 (Kolen & Brennan, 2004)。本研究採取水平等化的方式取得題目的鑑別度 (a)、難度 (b) 及猜測度 (c) 三參數的估計值。以XCALIBRE軟體之三參數模式估計IRT試題參數，包括：鑑別度參數、難度參數及猜測參數。最後再依據下述標準篩選題目：

1. 刪除IRT分析結果，試題參數接近軟體內定極限值之試題 ($a < 0.30$; $b > 2.95$ 或 $b < -2.95$; $c > 0.40$ 之情形)。
2. 刪除鑑別度 (D) 太低且點二系列相關太低之試題 (D 且 $r_{nb} < .2$)。
3. 刪除鑑別度太低且答對率太高或太低之試題 ($D < .2$ 且 $P < .2$ 或 $D < .2$ 且 $P > .9$)。

表7

GITC-G23之甲、乙、丙三式試題鑑別度、難度及猜測參數

分測驗	甲式		乙式		丙式	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
鑑別度參數 (a)						
語文反義	0.900	0.142	0.929	0.179	0.912	0.178
數字系列	1.052	0.149	1.001	0.135	1.089	0.156
圖形類比	0.975	0.189	0.964	0.141	1.021	0.140
難度參數 (b)						
語文反義	0.178	0.128	0.179	0.122	0.180	0.181
數字系列	0.166	0.679	0.169	0.636	0.168	0.649
圖形類比	0.099	0.893	0.092	0.907	0.094	0.966
猜測參數 (c)						
語文反義	0.214	0.009	0.212	0.011	0.214	0.018
數字系列	0.214	0.007	0.217	0.011	0.211	0.008
圖形類比	0.216	0.005	0.212	0.021	0.216	0.008

4. 參照IRT分析各題目之特徵曲線，找出難度適中且具良好鑑別度之題目。

針對各測驗題庫進行上述原則篩選後，再依IRT所得各題之難度參數 (b)，將各題庫之試題由易至難排列，再依S型分成甲、乙、丙三式正式題本。表7呈現三式測驗之IRT參數鑑別度 (a) 和難度 (b) 及猜測參數 (c)。由表7得知，甲式測驗的平均鑑別度為0.900 ~ 1.052之間，平均的難度介於0.099 ~ 0.178之間，平均的猜測度介於 .214 ~ 0.216；乙式測驗的平均鑑別度為 0.929 ~ 1.001之間，平均的難度介於 0.092 ~ 0.179之間，平均的猜測度介於 0.212 ~ 0.217；丙式測驗的平均鑑別度為 0.912 ~ 1.089之間，平均的難度介於0.094 ~ 0.180間，平均的猜測度介於 0.211 ~ 0.216。顯示三式測驗的鑑別度與難度相當。整體而言，本測驗的難度適中 ($-0.5 < b < 0.5$)，鑑別學生程度的能力尚佳 ($0.5 < a < 1$)，猜測度落在可接受的範圍 ($0.15 < c < 0.25$)。經由IRT的試題分析結果顯示，本研究所編製甲、乙、丙三式測驗均屬於適宜的試卷。

為了更精確的評估每式測驗所提供之訊息是否相當，本研究更進一步探討三式測驗在測驗特徵曲線（test characteristic curve, TCC）、測驗訊息曲線（test information function, TIF）及測驗標準誤曲線（test standard error, TSE）間的相關。TCC可以看出每個能力點的受試在全測驗所有試題的期望答對率，也可以將答對率乘上題數而得到理論的答對題數，從TCC的形狀也可以評鑑測驗對每一能力點的鑑別功能，曲線斜率越大的地方，其鑑別力就越大。TIF是該測驗對各潛在特質水準的受試能力估計所能提供之心理計量的訊息量，故TIF可看出該測驗對每一能力點測量之精確性，訊息量越大

代表測量越精確。TSE是訊息的平方根之倒數，訊息越大，標準誤就越小，訊息越大的題目或測驗對受試能力的評估越精確，估計誤差越小（de Ayala, 2009; Hambleton & Swaminathan, 1985）。圖1、圖2及圖3分別為三式測驗之特徵曲線、測驗訊息曲線及測驗標準誤曲線圖之比較，圖的橫軸代表能力（或是特質），這是無法直接觀察的潛在特質，可以從負無限大至正無限大，0代表能力等於平均水準的受試，-3至+3代表能力比平均數低三個標準差至比平均數高三個標準差之受試。從圖1至圖3可以發現三式測驗的TCC、TIC、TSE都非常接近。

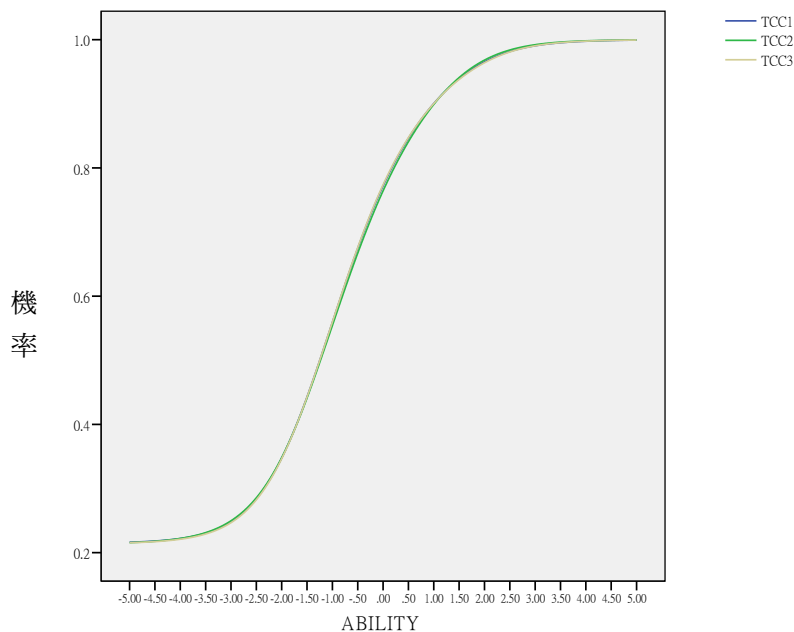


圖1 國小2-3年級團體智力測驗之測驗特徵曲線（TCC）

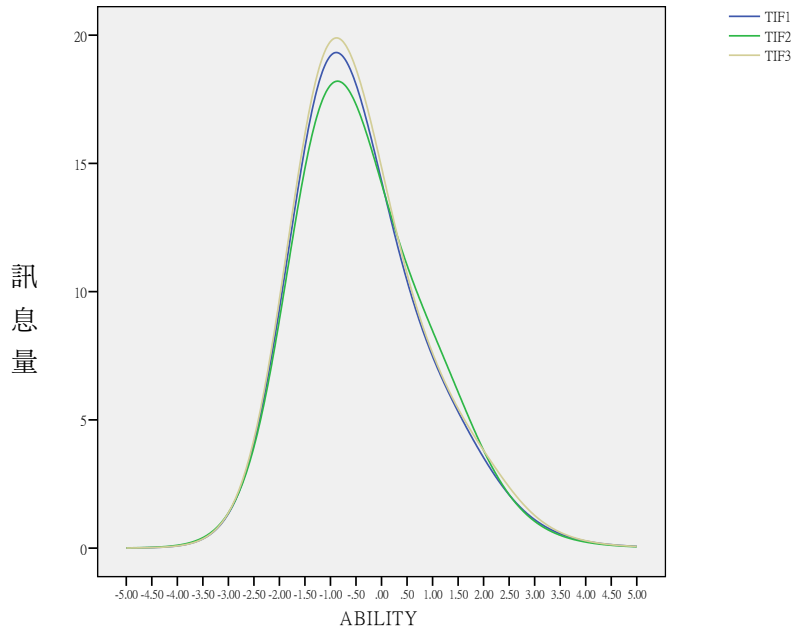


圖2 國小2-3年級團體智力測驗之測驗訊息曲線 (TIF)

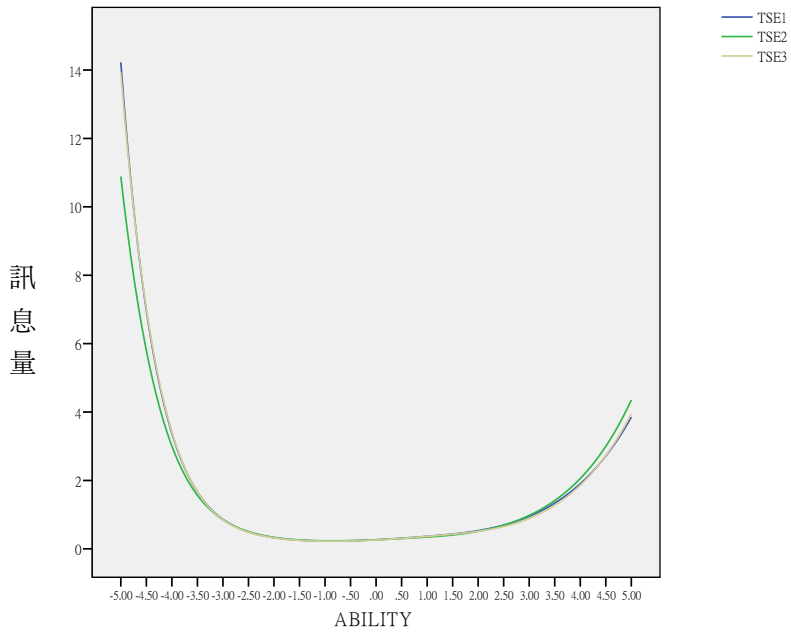


圖3 國小2-3年級團體智力測驗之測驗標準誤曲線 (TSE)

四、確定正式測驗題本

根據上述選題結果，最後，完成 GITC-G23正式題本各三式，每式各分成三個分測驗，其中「語文反義」分測驗有19題，「數字系列」分測驗有19題，「圖形類比」分測驗有21題。全測驗共有59題。本測驗包括三個分測驗，茲將本測驗之施測、及各分測驗內容簡述於後：

(一) 施測時間評估

為了解主試者施測流程與方式、時間評估及受試者的反應等情形。本正式測驗之施測時間評估共分為兩次，第一次施測以高雄市凱旋國小4年級一班學生進行施測時間評估，結果發現語文反義施測時間過長（9分鐘），於是將語文反義施測時間修改為8分鐘；數字系列施測時間則符合原先設定12分鐘；圖形類比施測時間過長（11分鐘），於是將圖形類比施測時間修改為10分鐘；第二次施測根據第一次修正之結果，以高雄市七賢國小2-3年級各兩班的學生為樣本，決定施測時間及施測流程；依據此次施測之受試者反應及時間估計，重新評估及修正各分測驗施測流程及所需時間。

(二) 正式測驗內容

本測驗需要受過測驗訓練的人員協助施測，以採團體施測方式為主。作答方式採五個選項，請受試者依據自己對各題的想法，分別就五個選項中擇其一作為答案。每一個分測驗均有時間限制，實際作答時間為30

分，施測說明約10分，一般受試者完成全測驗共需40分。正式測驗內容可說明如下：

1. 語文反義測驗

語文反義測驗的題目是給予適當語詞，請受試者由答案選項中找出相反的語詞，共有19題。

2. 數字系列

數字系列是一種數字的變化，數字間存在不同的變化關係，依照某一規則和順序，找出排列法則，從選項中找出適當的答案。共有19題，其內容可區分為下列兩種：（1）數字運算：數字的排列，可由四則運算計算出正確答案，為基礎四則運算題型。（2）數字推算：數字間排列存在某一種規則變化，受試者需推斷，方可推算出正確答案。

3. 圖形類比測驗

圖形類比測驗是由一組具體或抽象的圖形組成，題目的圖形提示中皆呈現某種規律的變化，藉由圖形的上方與下方、左方與右方、對角的變化規則，從選項中選出適當的答案，共有21題。內容可區分為圖形對照及圖形變化兩種類型：（1）圖形對照：可由圖形變化的規則，找出相對應的圖形。（2）圖形變化：圖形因增加或減少新屬性而改變；或各個圖形間的連續變化規則，依序增加或減少。

正式測驗之甲、乙、丙三式，依上述題型分類，各式之題型分佈之題數，如表8所示。

表8
正式題本各題型分佈之題數

分測驗	題型	甲式	乙式	丙式
語文反義	語文反義	19	19	19
數字系列	數字運算	12	12	12
	數字推算	7	7	7
圖形類比	圖形對照	12	8	8
	圖形變化	9	13	13

五、常模建立

在考量時間、人力及經費因素，本測驗目前僅建立甲式測驗之全國常模。參考教育部教育統計中有關國小2-3年級資料，顧及台灣地區北、中、南、東四區母群體樣本結構特徵的分佈狀況，採分層隨機叢集取樣方式取樣，以班級學生為單位進行分層隨機取樣，各區學校的抽取是先由電腦依據亂數抽取所需施測的學校名單，再經電話連繫取得同意。碰到無法協助的學校時就再重新進行電腦抽樣，共抽取18所國小，每校2-3年級各施測兩班，預計回收2,153人，最後回收2,135人。

表9

GITC-G23 (甲式) 之內部一致性Cronbach α 係數

年級	人數	語文反義	數字系列	圖形類比	總分
二年級	1,056	.778	.835	.795	.888
三年級	1,079	.712	.803	.752	.864
全體	2,135	.773	.823	.777	.882

(二) 穩定係數

甲式測驗以屏東市仁愛國小2-3年級各兩班共134名樣本，進行間隔四週之重測信度，三個分測驗的重測信度係數介於 .724至

表10

GITC-G23 (甲式) 之重測信度係數

年級	人數	語文反義	數字系列	圖形類比	總分
二年級	67	.845	.761	.724	.853
三年級	67	.885	.879	.753	.935
全體	134	.871	.813	.738	.893

(三) 複本信度係數

以高雄市新莊國小、中庄國小及鎮北國小2-3年級各三班，每班分別施測本測驗之三式分測驗，計算三式分測驗之複本信度係數。每式測驗各有171名學生填寫，但在語文反義有1名學生未填答，圖形類比分測驗有

肆、研究結果

本測驗編製甲、乙、丙三式乃根據IRT進行選題，並分析出三式測驗的測驗訊息相當一致。另本測驗之三式測驗亦具有不錯的複本信度。在進行本測驗的信度及效度考驗時，為節省施測人力及時間，本研究僅針對甲式測驗進行信度及效度考驗，茲說明如下：

一、信度考驗

(一) 內部一致性係數

依甲式測驗2,135名國小2-3年級學生常模樣本，採Cronbach α 係數計算各分測驗和全測驗的內部一致性係數，三個分測驗Cronbach α 係數介於 .712至 .835之間，全測驗則為 .882，請見表9。顯示本測驗之內部一致性良好。

.885之間，全測驗的重測信度係數為 .893，請見表10。所有的相關係數皆達 .01顯著水準，可見甲式測驗之穩定性頗佳。

8名學生未填答，故出現人數不一致情形。從表11發現語文反義測驗複本信度係數介於 .752至 .794之間；數字系列測驗介於 .725至 .779之間；圖形類比測驗介於 .735至 .760之間。顯示國小4-6年級團體智力測驗之各式測驗間的複本信度佳。

表11

GITC-G23複本信度係數

分測驗	甲式	乙式	丙式
語文反義			
甲式(N=170)	1.000		
乙式(N=170)	.794	1.000	
丙式(N=170)	.783	.752	1.000
數字系列			
甲式(N=171)	1.000		
乙式(N=171)	.725	1.000	
丙式(N=171)	.758	.779	1.000
圖形類比			
甲式(N=163)	1.000		
乙式(N=163)	.735	1.000	
丙式(N=163)	.748	.760	1.000

二、效度考驗

本測驗以內容效度、效標效度、及建構效度等三方面來驗證效度。

(一) 內容效度

本測驗題目皆由國小普通班、資優班資優教師協助命題，測驗編製考量智力理論：晶體智力及流體智力，量表內容參考過去智力測驗工具編製重點，例如：語文類測驗包括語文反義測驗（翁儷禎，1997），數學類測驗包括數字系列測驗（林幸台，2002；黃國彥等人，1983；翁儷禎，1997；Lohman & Hagen, 2001），非語文類測驗則包括圖形類比測驗（吳訓生，2007；吳武典等人，1996；林美珍等人，2002；俞筱鈞，1993；徐正穩、路君約，1995；許天威、蕭金土，1999；Brown, Sherbenou, & Johnson, 1997；Hammill et al., 1997；Lohman & Hagen, 2001；McCallum & Bracken, 1998；Naglieri, 2003）。最後，依語文、數學及圖形層面發展而成。另本測驗命題時建立雙向細目表，主要根據試題型式（語文反義、數字系列及、圖形類比）及認知能力（概念理解、推理思考）兩個向度進行題目編製，經由建立

雙向細目表可以幫助命題者釐清試題型式和認知能力內容的關係，以確保測驗能反映評量的重點。最後，測驗預試題本經由國小資深專家教師及學者專家討論與審查。因此，本測驗架構決定歷程、內涵、選取題項歷程，本測驗已力求提高內容效度。

(二) 效標效度

1. 與學業成績的相關

為了解甲式測驗表現與學校成績的相關，本研究以高雄市七賢國小、中正國小、及屏東市仁愛國小二年級各一班學生為樣本，計算月考成績與本測驗的相關。從表12可知甲式各分測驗及全測驗得分均與學校各科成績均達顯著差異，僅七賢國小學生之數字系列分測驗與國語成績相關未達顯著差異。語文反義與學校之各科的成績均達顯著相關介於 .441 ~ .608；數字系列與學校各科成績相關介於 .241 ~ .670；圖形類比則與學校各科成績多數達顯著相關介於 .304 ~ .578。另本測驗總分與學校各科成績介於 .503 ~ .739。顯示甲式測驗與學生在校成績的效標關聯效度佳。

表12

GITC-G23 (甲式) 與學業成績相關

學校	年級	學科	語文反義	數字系列	圖形類比	總分
七賢國小 (N=27)	二年級	國語	.552**	.241	.304*	.503**
		數學	.515**	.563**	.464**	.702**
中正國小 (N=28)	二年級	國語	.492**	.495**	.502**	.634**
		數學	.441**	.560**	.578**	.695**
仁愛國小 (N=30)	二年級	國語	.608**	.523**	.327**	.654**
		數學	.540**	.670**	.450**	.739**

* $p < .05$ ** $p < .01$

2. 與其他智力測驗的相關

甲式測驗以「國民小學語文智力測驗」(林幸台, 2002)、及「托尼非語文智力測驗」(吳武典等人, 1996)為效標, 探討本測驗與兩個效標測驗間之相關, 以原高雄縣中正國小二年級學生28名為樣本, 驗證本測驗之效標關聯效度。由表13顯示, 結果發現本測驗之語文反義測驗與「國民小學語文智力測驗」之各分測驗間的相關係數介於 .505

至 .670之間均達顯著相關; 數字系列測驗與「國民小學語文智力測驗」各分測驗間的相關係數介於 .479至 .668之間, 均達顯著相關; 圖形類比與「國民小學語文智力測驗」各分測驗間的相關係數介於 .133至 .427之間, 部分達顯著相關。另本測驗之圖形類比測驗與「托尼非語文智力測驗」之相關係數為 .729, 達顯著相關。結果顯示甲式測驗的效標關聯效度佳。

表13

GITC-G23 (甲式) 得分與其他智力測驗得分之相關 (N=28)

	語文反義	數字系列	圖形類比	總分
國民小學語文智力測驗				
歸類測驗	.633**	.588**	.261	.605**
數的比較	.505*	.521*	.133	.434*
替代比較	.585**	.668**	.384*	.665**
語文類推	.670**	.479*	.427*	.616**
托尼非語文智力測驗	.055	.317	.729**	.461*

* $p < .05$, ** $p < .10$

3. 不同認知能力學生之區別力

以高雄市陽明國小、新莊國小及福康國小一般智能資優班學生為對象, 國小二年級學生共56名, 三年級學生共70名, 其中國小二年級資優學生來源, 乃取自二年級下學期通過高雄市國小資優鑑定資格, 即將國小三年進入資優班就讀的學生, 以探討一般智能資優班學生在甲式測驗得分與常模樣本得分的差異情形。從表14得知, 本測驗語文反義、數字系列及圖形類比三個分測驗及全

測驗得分均顯著高於常模樣本得分, 由母群標準差已知單一樣本Z檢定所得到的Z值均在2.33以上, 結果顯示本測驗均可有效區別一般智能資優生與普通生。此外, 上述小二(56名)、小三(70名)資優學生在甲式測驗總分對照常模樣本總分後, 結果發現小二、小三資優學生高於正兩個標準差以上人數比率分別為96%、94%, 顯示本測驗可有效區別資優學生與普通學生。

表14
資優生與常模樣本在GITC-G23（甲式）各分測驗之差異比較

分測驗	二年級				Z	三年級				
	資優生(N=56)		常模(N=1,056)			資優生(N=70)		常模(N=1,079)		Z
	M	SD	M	SD		M	SD	M	SD	
語文反義	17.04	1.17	13.38	3.49	7.85**	17.96	1.06	15.57	2.75	7.26**
數字系列	18.29	1.02	14.65	3.79	7.19**	18.70	0.52	16.17	3.07	6.89**
圖形類比	20.11	0.93	17.19	3.40	6.65**	20.36	0.89	17.77	2.87	7.55**
總分	55.43	2.05	45.23	8.58	8.67**	57.01	1.45	49.51	6.97	9.00**

** $p < .01$

(三) 建構效度

1. 分測驗間相關

表15顯示甲式測驗常模樣本各分測驗之間，以及各分測驗與總分的相關，結果發

表15

GITC-G23（甲式）各分測驗之相關

分測驗	語文反義	數字系列	圖形類比	總分
語文反義	1.000			
數字系列	.514**	1.000		
圖形類比	.444**	.486**	1.000	
總分	.807**	.836**	.783**	1.000

** $p < .01$

現語文反義、數字系列及圖形類比分測驗之間的相關介於 .444至 .514之間，均屬中度相關；而各分測驗與總分之相關則在 .783至 .836之間，均呈現高度相關。

2. 驗證性因素分析

為驗證甲式測驗的內在因素架構，本研究以常模樣本的資料進行驗證性因素分析。依甲式各分測驗內容之題型區分：語文反義皆為找出相反的語詞，故為一個量表；數字系列區分為數字運算及數字推算兩個量表；圖形類比區分為圖形對照及圖形變化兩個量表。全測驗共分為五個量表。根據文獻探討，本測驗第一個因素為晶體智力，是由語文反義、數字運算、數字推算所組成；第二個因素為流體智力，是由圖形對照及圖形變化所組成。研究者參考國外著名智力測驗採用驗證性因素分析之作法，如魏氏兒童智力量表第四版（Canivez, 2014; Devena, Gay & Watkins, 2013; Nakano & Watkins, 2013;

Watkins, 2010）及魏氏成人智力量表第四版（Canivez & Watkins, 2010），其進行智力測驗之驗證性因素分析，亦納入部分潛在變項只有涵蓋兩個測量指標。雖然本測驗流體智力因素僅能分成圖形對照及圖形變化兩個測量指標，應不影響驗證性因素分析的結果。故本研究以結構方程模式適合度評鑑進行初階驗證性因素分析，來檢驗GITC-G23（甲式）的建構效度，以常模有效樣本國小2-3年級共2135名學生進行多向度初階因素模式的適配度考驗，採用AMOS 20.0版統計軟體進行資料分析，以最大概似可能法（maximum likelihood method, ML）進行參數估計。檢驗的模式圖與結果如圖4所示：

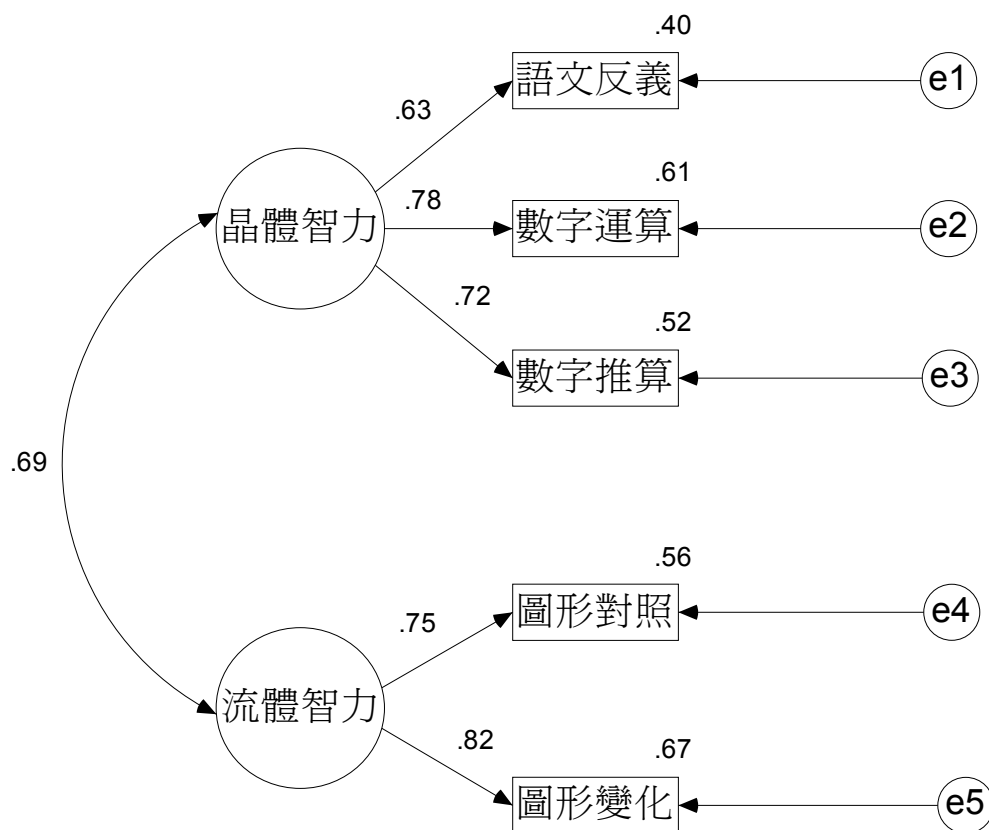


圖4 GITC-G23 (甲式) 多向度初階因素模式圖 (標準化解)

由表16得知，在模式基本的適合標準部分：模式沒有負的誤差變異，誤差變異皆達顯著水準，參數間相關的絕對值都未太接近1，因素負荷量大都介於.5-.95之間。在整體模式適合度部分：本研究分析模式與觀察資料適配的卡方考驗為 ($p < .001$)，此表示所提出分析模式並沒有與觀察資料適配。然而，卡方檢定對樣本數相當敏感，Hair、Black、Babin、Anderson與Tatham (2006) 指出，當樣本數大於400人時，很多模式都可能被拒絕，本研究的分析樣本高達2000人，值容易因大樣本之影響而顯著，造成模式的不適配，或許是理論模式被拒絕之故，不適合作為評斷的指標。因此採用NFI、GFI、RFI、IFI、TLI、CFI等比較不會受樣本影響

的指標來評估模式的適合度，由表16多向度初階模式適合度指標可知，標準化SRMR為.026 (小於.05)，GFI、AGFI、NFI、IFI、TLI皆大於.95，RMSEA值為.074，雖大於.05，但小於.08，為模式契合度可以接受的門檻 (McDonald & Ho, 2002)。另根據邱皓政 (2003) 論點，當必須進行不同競爭模式之比較，或從多個模式挑選一最佳模式時，ECVI值 (期望的交互驗證指數) 可以做為模式選擇的依據。一般而言，ECVI值愈小表示模式愈簡約，本測驗ECVI值為.580小於獨立模式之ECVI值 (1.53)。綜上發現本模式適合度良好，顯示本測驗因素架構獲得驗證。

表16
GITC-G23 (甲式) 因素分析負荷量 (N=2,135)

評鑑項目	實際數值	評鑑結果
基本適合標準		
是否沒有負的誤差變異?	是	良好
誤差變異是否都達顯著水準?	是	良好
參數間相關的絕對值是否未太接近1?	是	良好
因素負荷量是否介於.5-.95之間?	介於.63-.82之間	良好
整體與比較模式適合標準		
χ^2 值是否未達顯著?	$\chi^2=50.967$	因卡方值會受樣本數影響，所以僅做為參考值
GFI指數是否大於.95?	.990	良好
AGFI指數是否大於.95?	.964	良好
SRMR指數是否低於.05?	.026	良好
RMSEA是否小於.05?	.074	尚可
χ^2 值比率是否小於3?	12.742	因卡方值會受樣本數影響，所以僅做為參考值
Δ_1 (NFI) 指數是否大於.95?	.984	良好
Δ_2 (IFI) 指數是否大於.95?	.986	良好
TLI (NNFI) 指數是否大於.95?	.964	良好
ECVI愈小愈好?	.580	良好

三、常模建置

(一) 常模樣本測驗分數之平均數及標準差

甲式測驗之常模建置乃參考教育部教育統計資料，顧及母群體樣本結構特徵的分佈狀況，採分層隨機叢集取樣方式取樣，以班

級學生為單位進行分層隨機取樣，計18所學校，每校2-3年級各施測兩班，共取得常模樣本為2,135名，關於常模樣本人數分配表，請見表17。本研究常模樣本在各分測驗上之得分平均數及標準差如表18所示。

表17
常模樣本基本資料

區域	縣市	學校	人數	年級	男	女	合計		
北區	台北市	民生國小	133	二	37	29	66		
				三	34	33	67		
	新北市	關渡國小	124	二	30	31	61		
				三	35	28	63		
		豐年國小	118	二	32	23	55		
				三	31	32	63		
		義學國小	127	二	30	28	58		
				三	39	30	69		
	桃園縣	中壢國小	105	二	30	25	55		
				三	31	19	50		
		楊梅國小	122	二	33	24	57		
	中區	台中市	立人國小	115	二	35	25	60	
三					29	26	55		
春安國小					82	二	23	20	43
彰化縣			平和國小	134	二	35	28	63	
					三	38	33	71	
			花壇國小	127	二	35	29	64	
南區		高雄市	七賢國小	117	二	35	25	60	
					三	25	32	57	
					漢民國小	125	二	33	29
		台南市	勝利國小	135	二	37	31	68	
					三	36	31	67	
			新南國小	105	二	31	22	53	
屏東縣	仁愛國小	138	二	36	32	68			
			三	35	35	70			
	內埔國小	131	二	36	30	66			
東區	台東市	東大附小	125	二	37	30	67		
				三	33	25	58		
				太平國小	72	二	14	16	30
						三	20	22	42
總計					1,004	1,131	2,135		

表18
GITC-G23 (甲式) 各年級的平均數和標準差

測驗	二年級			三年級			總計		
	男	女	合計	男	女	合計	男	女	合計
取樣人數	579	477	1,056	552	527	1,079	1,131	1,004	2,135
語文反義	<i>M</i> 13.15	13.66	13.38	15.47	15.67	15.57	14.28	14.72	14.49
	<i>SD</i> 3.65	3.25	3.49	2.73	2.76	2.75	3.44	3.16	3.32
數字系列	<i>M</i> 14.73	14.56	14.65	16.23	16.10	16.17	15.47	15.37	15.42
	<i>SD</i> 3.83	3.75	3.79	3.11	3.02	3.07	3.58	3.47	3.53
圖形類比	<i>M</i> 17.11	17.30	17.19	17.88	17.67	17.77	17.48	17.49	17.49
	<i>SD</i> 3.52	3.24	3.40	2.75	3.00	2.87	3.19	3.12	3.16
測驗總分	<i>M</i> 44.99	45.52	45.23	49.58	49.43	49.51	47.23	47.57	47.39
	<i>SD</i> 8.93	8.13	8.58	6.91	7.05	6.97	8.33	7.83	8.10

(二) 不同年級、性別樣本在各分測驗上
得分之平均數差異考驗

為了解不同年級、性別樣本在各分測驗上得分是否有顯著差異，以作為建立常模之依據。乃以年級、性別作為自變項進行二因子變異數分析，結果如表19所示，各年級的平均數間有顯著差異，在國小2-3年級部分，三個分測驗得分均是三年級高於二年級。而

性別之間，國小2-3年級部分，除語文反義有差異外，數字系列及圖形類比並未出現顯著差異，依性別方面，女生在語文反義明顯優於男生；在數字系列及圖形類比分測驗之男生與女生沒有顯著差異。從全測驗得分來看，年級有達到顯著差異，而男女生的總分沒有顯著差異，最後，僅針對年級建立不同常模，並未建立性別常模。

表19
GITC-G23 (甲式) 常模樣本施測結果變異數分析摘要

分量表	變異來源	<i>df</i>	<i>MS</i>	<i>F</i> 值	差異考驗
語文反義	年級	1	2490.349	254.636**	3>2
	性別	1	69.527	7.109	女>男
	年級X性別	1	12.407	1.269	
數字系列	年級	1	1230.925	103.618**	3>2
	性別	1	12.104	1.019	
	年級X性別	1	.492	.041	
圖形類比	年級	1	180.573	18.352**	3>2
	性別	1	.093	.009	
	年級X性別	1	22.105	2.247	
總分	年級	1	9687.616	158.975**	3>2
	性別	1	20.736	.340	
	年級X性別	1	56.589	.929	

***p* < .01

伍、討論與建議

一、研究結果討論

本研究目的在編製三式GITC-G23適用於評量國小學生之團體智力測驗，以提供資優學生鑑定之用。本測驗以評量智力重要元素：晶體與流體智力為基礎，測驗內容兼具語文與非語文，以求節省施測時間，並編製多套複本測驗，以減少題目外流。另測驗編製歷程應用試題反應理論，使本測驗組題更具嚴謹性。GITC-G23包含語文反義、數字系列與圖形類比等三個分測驗，並以台灣地區2,135名國小2-3年級學生為常模樣本，建立不同年級的常模。整體而言，本測驗具有不錯的信度及效度，可供現場教育使用。研究結果顯示，在信度方面，本測驗內部一致性係數，在國小2-3年級團體智力測驗之三個分測驗Cronbach α 係數介於.712至.835之間，全測驗為.882；分測驗四週重測信度係數，三個分測驗的重測信度係數介於.724至.885之間，全測驗為.893；三式測驗複本信度係數則介於.725至.794之間。根據Kaplan與Saccuzzo (2013)指出，當測驗信度在.70以上時，顯示研究具有相當可靠的信度。本研究顯示國小2-3年級團體智力測驗具有不錯的內部一致性、重測信度及複本信度。

就效度而言，本研究除具內容效度外，本團體測驗分別與學生在校的學業成績有達至中度以上相關，此結果與吳武典等人(2010)發現相符。另本測驗與「國民小學語文智力測驗」及「托尼非語文智力測驗」亦有不錯的效標關聯效度。整體而言，本測驗與效標測驗具有不錯相關，顯示有較佳的效標關聯效度。本測驗亦可區別資優生與普通生之差異，顯見具有不錯區別力。此結果發現與王金香和蕭金土(2006)研究相符，團體性向測驗能有效區別一般生與資優生。本研究結果亦與Rizza、McIntosh 和 McCunn

(2001)研究發現相同，其曾探討51名資優生與51名普通生在WJ-III的測驗表現，結果發現資優生在Gf與Gc能力表現優於普通學生。驗證性因素分析結果顯示與原先建構測驗因素相符合，而且結構相當穩定。本研究得到因素架構與過去文獻結果相符，如吳訓生(2007)編製國民小學學業性向測驗亦得到兩個因素，第一個因素屬語文與數字符號之推理能力，第二個因素屬圖形與空間關係之推理能力。另吳武典等人(2010)及Otis與Lennon(1996)編製團體智力測驗內容可得語文、非語文兩因素，均與本研所得晶體智力、流體智力因素相符。

本研究為了解不同年級、性別樣本在各分測驗上得分是否有顯著差異，以年級、性別作為自變項進行二因子變異數分析，在年級方面，結果發現不同年級學生在三個分測驗均達顯著差異，三個分測驗得分皆為三年級高於二年級，顯示年齡愈高，智力發展愈佳，此與過去文獻發現相一致(吳武典等人, 2010; 吳訓生, 2007; 林幸台, 2002; 簡茂發, 1981)。在性別方面，除語文反義有差異外，數字系列及圖形類比並未出現顯著差異，其中女生在語文反義明顯優於男生；在數字系列及圖形類比男生與女生沒有顯著差異。此性別差異之結果，與吳訓生(2007)、林幸台(2002)、路君約等人(2001)研究結果類似。未來在使用本測驗時，宜留意不同年級在各分測驗出現的差異，以及不同性別在語文反義分測驗表現之差異。

二、研究建議

(一)繼續驗證乙、丙兩式測驗的信度及效度

本研究雖編製三式國小2-3年級團體智力測驗，但由於常模建立耗費相當多資源。研究者根據所有題目，採用IRT方法，依各

分測驗之題目難易度區分成三式，盡量使三式測驗難度及鑑別度相當，且經由複本信度考驗後也發現三式測驗之各分測驗間，具有不錯複本信度。最後考量人力、時間與經費下，故最後僅以本測驗之甲式測驗進行信度、效度考驗，以及建立甲式測驗之常模。建議未來可進一步針對乙、丙兩式測驗進行信度、效度驗證，並新增建立乙、丙兩式之常模。

(二) 針對不同背景變項學生進行分析

由於本研究針對不同背景變項僅分析學生之性別及年級差異，因此對於不同地區（如都市、鄉鎮區域）、學生出生序等，或家庭變項，如家庭結構、雙親教育程度等情形，仍有待更深入探討。

(三) 驗證本測驗運用在資優學生鑑定的合適性

本測驗甲式雖已建立全國常模資料，但在國小資優學生的鑑定，除了團體智力測驗的資料外，還需要個別智力測驗...等方面資料進行綜合研判，不宜單獨依據本測驗統計標準作決定。未來可進一步探討學生在本團體智力測驗得分與通過個別智力測驗者之差異性，考驗本測驗在資優學生鑑定的適切性，例如：Beal (1996) 曾探討團體智力測驗與個別智力測驗運用在資優學生鑑定之差異性。故本測驗在運用於資優學生鑑定的合適性仍需要未來研究的考驗。

致謝

本文為研究者所主持教育部專案計畫之部份成果改寫而成，本計畫承吳裕益教授及郭靜姿教授協助，以及黃慧慈助理協助資料分析，特此致謝。

參考文獻

一、中文部分

- 王金香、蕭金土 (2006)。團體性向測驗在資賦優異學生甄試運用之研究。*教育研究與發展期刊*, 2 (4), 143-162。
- 朱錦鳳 (2005)。圖形思考智能測驗。臺北市：心理。
- 吳武典、金瑜、張靖卿 (2010)。多向度團體智力測驗 (兒童版)。臺北市：心理。
- 吳武典、蔡崇建、胡心慈、王振德、林幸台、郭靜姿 (1996)。托尼非語文智力測驗指導手冊。臺北市：心理。
- 吳訓生 (2007)。國民小學學業性向測驗指導手冊。臺北市：教育部特殊教育工作小組。
- 林幸台 (2002)。國民小學團體語文智力測驗指導手冊。臺北市：教育部特殊教育工作小組。
- 林美珍、林邦傑、黃國彥、李富言 (2002)。國民小學團體非語文智力測驗指導手冊。臺北市：教育部特殊教育工作小組。
- 邱皓政 (2003)。結構方程模式。臺北市：雙葉。
- 俞筱鈞 (1993)。瑞文氏圖形推理測驗系列指導手冊。臺北市：中國行為科學社。
- 徐正穩、路君約 (1995)。圖形式智力測驗指導手冊。臺北市：中國行為科學社。
- 許天威、蕭金土 (1999)。綜合性非語文智力測驗。臺北市：心理。
- 黃國彥、鍾思嘉、傅粹馨 (1977)。羅桑二氏非語文智力測驗指導手冊。臺北市：正昇教育科學社。
- 黃國彥、鍾思嘉、林珊如、李良哲 (1983)。羅桑二氏語文智力測驗指導手冊。臺北市：正昇教育科學社。
- 陳榮華 (1998)。陳氏非語文能力測驗。臺北市：中國行為科學社。
- 陳龍安 (1996)。智能結構學習能力測驗。臺北市：心理。

- 翁儷禎 (1997)。推理思考測驗指導手冊。臺北市：測驗出版社。
- 教育部 (2008)。資優教育白皮書。臺北市：作者。
- 路君約、吳武典、簡明建 (2001)。國中學業性向測驗。臺北市：心理
- 路君約、盧欽銘 (1991)。國民中學智力測驗 (第三種)。臺北市：中國行為科學社。
- 簡茂發 (1981)。我國國小及國中學生的智力發展。教育心理學報，14，125-148。

二、英文部分

- Bracken, B. A., & Naglieri, J. A. (2003). Assessing diverse populations with nonverbal tests of general intelligence. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological & educational assessment of children: Intelligence, aptitude, and achievement* (pp. 243-274). New York, NY: The Guilford Press.
- Beal, A. L. (1996). A comparison of WISC-III and OLSAT-6 for the identification of gifted students. *Canadian Journal of School Psychology, 11*(2), 120-129.
- Brown, L., Sherbenou, R., & Johnson, S. (1997). *Test of Nonverbal Intelligence* (3rd ed.). Austin, TX: PRO-ED
- Canivez, G. L. (2014). Construct validity of the WISC-IV with a referred sample: Direct versus indirect hierarchical structures. *School Psychology Quarterly, 29*(1), 38-51.
- Canivez, G. L., & Watkins, M. W. (2010). Investigation of the factor structure of the Wechsler Adult Intelligence-fourth edition (WAIS-IV): Exploratory and higher order factor analysis. *Psychological Assessment, 22*(4), 827-836.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press.
- Carroll, J. B. (2005). The three-stratum theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 69-76). New York, NY: The Guilford Press.
- Clark, B. (2012). *Growing up gifted: Developing the potential of children at school and home* (8th ed.). Upper Saddle River, NJ: Prentice Hall.
- Cooper, C. (1999). *Intelligence and abilities*. London, UK: Routledge.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford.
- Demetriou, A., & Efklides, A. (1994). Intelligence, mind and reasoning: Three levels of description. In A. Demetriou. & A. Efklides (Eds.), *Intelligence, mind, and reasoning: Structure and development* (pp. 1-7). Amsterdam, Holland: North-Holland.
- Devena, S. E., Gay, C. E., & Watkins, M. W. (2013). Confirmatory factor analysis of the WISC-IV in a hospital referral sample. *Journal of Psychoeducational Assessment, 31*(6), 591-599.
- Elliott, C. D. (2007). *Differential ability scales, 2nd edition: Introductory and technical handbook*. San Antonio, TX: The Psychological Corporation.
- Flanagan, D. P., & McGrew, K. S. (1997). A cross-battery approach to assessing and interpreting cognitive abilities: Narrowing the gap between practice and cognitive science. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 314-325). New York, NY: The Guilford Press.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer-Nijhoff Publishing.
- Hammill, D. D., Pearson, N. A., & Wiederholt, J. E. (1997). *Comprehensive Test of Nonverbal Intelligence*. Austin, TX: Pro-Ed.
- Horn, J. L., & Blankson, N. (2005). Foundations for better understanding of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 41-68). New York, NY: The Guilford Press.
- Hussain, L., Jamil, A., Siraji, M. J., & Maroof, K. (2012). Development and standardization of intelligence test for children. *International Journal of Learning & Development*, 2(5), 190-202.
- Kamphaus, R. W. (1993). *Clinical assessment of children's intelligence*. Boston, MA: Allyn and Bacon.
- Kamphaus, R. W., Petoskey, M. D., & Morgan, A. W. (1997). A history of intelligence test interpretation. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 17-31). New York, NY: The Guilford Press.
- Kaplan, R. M., & Saccuzzo D. P. (2013). *Psychological testing: Principles, applications, and issues* (8th ed.). Belmont, CA: Cengage Learning.
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman assessment battery for children* (2nd ed.). Circle Pines, MN: American Guidance Service.
- Kline, P. (1991). *Intelligence: The psychometric view*. London, UK: Routledge.
- Kolen, M. J., Brennan, R. L., (2004). *Test equating, scaling, and linking: Methods and practices*. New York, NY: Springer-Verlag.
- Lohman, D. F., & Hagen, E. (2001). *Cognitive Abilities Test*. Itasca, IL: Riverside.
- Maller, S. J. (2003). Best practices in detecting bias in nonverbal tests. In R. S. McCallum(Ed.), *Handbook of nonverbal assessment*(pp. 23-47). New York, NY: Kluwer Academic/Plenum Publishers.
- McCallum, R. S. (2003). Context for nonverbal assessment of intelligence and related abilities. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (pp. 3-21). New York, NY: Kluwer Academic/Plenum Publishers.
- McCallum, R. S., & Bracken, B. A. (1998). *The Universal Nonverbal Intelligence Test*. Itasca, IL: Riverside.
- McDonald, R. P., & Ho, M. H. R. (2002). Principles and Practice in Reporting Statistical Equation Analyses. *Psychological Methods*, 7 (1), 64-82.
- Murphy, K. R., & Davidshofer, C. O. (2004). *Psychological testing: Principles and applications* (6th ed.). Upper Saddle River, NJ: Prentice Hall.
- Naglieri, J. A. (2003). *Naglieri nonverbal ability test*. San Antonio, TX: The Psychological Corporation.
- Naglieri, J. A., & Bardos, A. N. (1997). *Manual for the General Abilities Test for Adults*. Minneapolis, MN: National Computer Systems.
- Naglieri, J. A., & Das, J. P. (1997). *Das-Naglieri: Cognitive assessment system*. Itasca, IL: Riverside Publishing.
- Nakano, S., & Watkins, M. W. (2013). Factor structure of the Wechsler Intelligence Scale for Children-fourth edition among referred native American students. *Psychology in the Schools*, 50(10), 957-968.
- Otis, A., & Lennon, R. (1996). *Otis-Lennon school ability test* (7th ed.). San Antonio, TX: Harcourt Brace Educational Measurement.

- Preuss, L. J., & Dubow, E. F. (2004). A comparison between intellectually gifted and typical children in their coping responses to a school and a peer stressor. *Roepers Review*, 26(2), 105-111.
- Rizza, M. G., McIntosh, D. E., & McCunn, A. (2001). Profile analysis of the Woodcock-Johnson III tests of cognitive abilities with gifted students. *Psychology in the Schools*, 38(5), 447-455.
- Roid, G. H. (2003). *Stanford-Binet Intelligence Scales* (5th ed.). Itasca, IL: Riverside Publishing.
- Stankov, L. (1991). The effects of training and practices on human abilities. In A. H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement* (pp. 97-118). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York, NY: Cambridge University Press.
- Sternberg, R. J. (1992). Ability tests, measurements, and markets. *Journal of Educational Psychology*, 84, 134-140.
- Swanson, H. L. (1985). Assessing learning disabled children's intellectual performance: An information processing perspective. In K. D. Gadow (Ed.), *Advances in learning and behavioral disabilities* (pp. 225-272). Greenwich, CT: JAI Press.
- Volker, M. A., Lopata, C., & Cook-Cottone, C. (2006). Assessment of children with intellectual giftedness and reading disabilities. *Psychology in the Schools*, 43(8), 855-869.
- Watkins, M. W. (2010). Structure of the Wechsler Intelligence Scale for Children-fourth edition among a national sample of referred students. *Psychological Assessment*, 22(4), 782-787.
- Wechsler, D. (2003). *The WISC-IV technical and interpretive manual*. San Antonio, TX: Psychological Corporation.
- Wilhoit, B. E., & McCallum, R. S. (2003). Cross-battery assessment of nonverbal cognitive ability. In R. S. McCallum (ed.), *Handbook of nonverbal assessment* (pp. 63-86). New York, NY: Kluwer Academic/Plenum Publishers.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock Johnson III tests of cognitive abilities*. Rolling Meadows, IL: Riverside Publishing.

The Development of a Group Intelligence Test for Children in Grades 2-3: A Reliability and Validity Study

Ming- Fu Tsai

Department of Special Education, Kaohsiung Normal University

Abstract

The purpose of this study was to develop a group intelligence test for children in grades 2-3, which could be used as an identification instrument for gifted students. The four major characteristics of the test were a theoretical basis on crystallized and fluid intelligences, a combination of verbal and nonverbal contents, parsimonious test time, and multiple test forms. It consisted of three subtests: Verbal Reasoning, Arithmetic Reasoning, and Figural Reasoning. It was a group test with a 40-minute time limit for each version. Results revealed that the Cronbach α coefficients of subscales were .712~.835, while for the overall test it was .882. The test-retest reliability coefficients over a period of four weeks were .724~.885, while for the overall test it was .893. The correlations between the different test versions were .725~.794. It was found that the total scores of the test were significantly related to the student's academic performance, the scores of a well-known verbal intelligence test and a nonverbal intelligence test. In addition, a confirmatory factor analysis yielded two factors (crystallized and fluid intelligences) for this test. Generally speaking, the group intelligence test for children (grades 2-3) exhibited acceptable indicator values for reliability and validity. To this end, the normalized t-score norms and percentile ranks for 2,135 subjects in the test (Form A) were also developed. Finally, some suggestions were proposed for future research.

Keywords: gifted student, group intelligence test, reliability, validity