
標準化測驗的選擇與解釋

新北市秀山特殊教育資源中心

主任 黃俊榮

於**25**期心評人員研習

內容大綱

壹 評量基本概念

貳 鑑定基準、構念與評量方式

參 標準化測驗的選擇與解釋

評量

❖ 做與學生有關的決定時，依**決策目的**而進行的資料收集過程。

—Salvia & Ysseldyke

不同決策目的，不同評量重點與方式

❖ 不同決策目的

- 篩選：全面性發現未覺察的問題
- 監控進步情形：學生是否適切的進步
- 調整教學計畫：教什麼？如何教？合理期待為何？
- 分配資源：學生是否需要資源與支持以達有效學習
- 方案執行成效：補救教學/學習扶助成果
- 績效責任：是否達到期望的成果
- 確認是否符合特殊教育資格（鑑定）
 - ▶ 達到障礙基準
 - ▶ 有特殊學習需求

測驗是評量（Assessment）方式之一

❖ 文件資料檢視與分析

❖ 晤談

❖ 觀察

❖ 工作/作業（**task**）樣本實作

❖ 測驗

- 標準化測驗：標準化程序+信效度檢驗+常模
- 自編測驗
 - ▶ 測得對，測到重點
 - ▶ 如何反映（解釋）學生的表現

執行心評工作時，測驗有其限制

- ❖ 心評時，需靠評量資料做決定和給教學建議
 - 特教資格與障別
 - 教育安置方式
 - 相關支持服務
 - 符應問題與需求的教學建議
- ❖ 單靠標準化測驗分數，無法完成這些決定
- ❖ 單憑身心障礙證明/醫療診斷證明，以不能完成這些決定。

以鑑定為目的的評量

❖ 鑑定時，通常會遇到兩種個案

- 已經接觸過，教學過一段時間的個案
 - ▶ 決定要提鑑定確認資格前，你已經知道什麼？有什麼資料？
 - ▶ 接著要做什麼？
- 沒有深入接觸的個案：
 - ▶ 從前一教育階段轉銜
 - ▶ 導師或家長要申請鑑定取得資格
 - ▶ 你手上有什麼資料？
 - ▶ 接著要做什麼？

想想，新北市為何強調學生輔導機制

不同收集資料取徑（approach）

❖ 全方位收集 vs. 聚焦問題收集

- 先廣泛收集資料，再根據初步分析後的結果，決定是否再收集其他資料。
- 你已經初判個案是哪種障礙，依據該障別的鑑定基準收集資料。

❖ 哪一種取徑適合你？

- 接觸過哪些學生的經驗、對障礙特質的正確理解程度。

❖ 不管走哪條路，小心「亡鈇意鄰」效應

鑑定基準與評量

❖ 決定特殊教育資格必須依據障礙的鑑定基準

❖ 鑑定辦法對各障礙通常有：

- 概念性定義 ➡ 成因與障礙表現描述
- 操作性標準（基準） ➡ 構成要件、符合障礙的標準

❖ 鑑定基準在說什麼？

- 指出應該評量的「構念」
- 指出要符合哪些條件，多嚴重才稱為「障礙」

鑑定基準與評量

❖ 難在：

- 這些構念是什麼意思？
 - ▶ 智力、閱讀理解、視野、知覺動作、注意力、情緒、適應行為、社會互動、溝通...
- 這些構念要怎麼評量？怎麼收集資料？
- 問題、困難到障礙是一條連續線；心評需要做決定，並說明這個決定是合理的。
 - ▶ 多嚴重、多顯著困難、表現多差才稱為障礙
- 請記得，基準是人定的：
 - ▶ 如智能障礙定義、自閉症定義的演變

從常見的例子開始—智能障礙鑑定

❖ 智能障礙的鑑定基準

- 本法第三條第一款所稱智能障礙，指個人之智能發展較同年齡者明顯遲緩，且在學習及生活適應能力表現上有顯著困難者。
- 前項所定智能障礙，其鑑定基準依下列各款規定：
 - ▶ 心智功能明顯低下或個別智力測驗結果未達平均數負二個標準差。
 - ▶ 學生在生活自理、動作與行動能力、語言與溝通、社會人際與情緒行為等任一向度及學科（領域）學習之表現較同年齡者有顯著困難情形。

從常見的例子開始—智能障礙鑑定

❖ 智能障礙

- 核心構念是什麼？
- 達障礙（明顯遲緩、顯著困難）的標準
- 評量方法、應收集的資料
- 注意事項

從常見的例子開始—智能障礙鑑定

❖ 對於智能障礙的鑑定

- 什麼的孩子你會想走智能障礙鑑定這條路？
- 醫學診斷、身心障礙證明/手冊和學生實際表現的一致性
- 如果沒有證明，鑑定標準說什麼？
 - ▶ 心智功能明顯低下的證據
 - ▶ 實施標準化智力測驗以及評估適應行為表現

❖ 選擇哪一種智力測驗？

- 這是智能障礙的鑑定，所以要「標準化個別智力測驗」。
- 單一向度、團體的測驗不適合。

從常見的例子開始—智能障礙鑑定

❖ 常用標準化個別智力測驗

- 魏氏兒童智力量表第四版 (WISC-IV)
- 魏氏兒童智力量表第五版 (WISC-V)
- 綜合心理能力測驗 (CMAS)
- 魏氏幼兒智力量表修訂版 (WPPSI-R)
- 魏氏幼兒智力量表第四版 (WPPSI-IV) (本市無)

❖ 需要受訓且領有證書才能使用

從常見的例子開始—智能障礙鑑定

❖ 這些測驗不建議用在「智能障礙鑑定」

- 托尼非語文智力測驗 (TONI-?)
- 瑞文氏矩陣推理測驗 (彩色、標準)
- 修訂畢保德圖畫詞彙測驗 (PPVT-R)
- 簡易個別智力量表

❖ 不建議使用的理由

- 測量的構念單一
- 常模的新近性

❖ 可以使用的時機

- 「估測」大致的智力表現

從常見的例子開始—智能障礙鑑定

❖ 使用智力測驗的幾個常見問題

- 以為個別施測，就是個別智力測驗
- 用低一層次的測驗推翻高一層測驗的結果
 - ▶ 篩選測驗、單一向度測驗 vs. 診斷測驗、多向度測驗
- 同一測驗在不同時期的測驗結果要如何解釋
- 相似但不同測驗表現如何解釋
- 該採全量表還是單一因素指數解釋智力？
- 分數剛好落在切截點上

❖ 遇到無法施測的學生怎麼辦？

- 釐清是障礙很嚴重？不配合？還是反應受限？
- 調整反應方式（但可能影響效度）
- 只好多元評量了

從常見的例子開始—智能障礙鑑定

❖ 年齡較高、智力稍高的智能障礙者之鑑定

- 智力接近臨界值
- 有些因素指數（或加上信賴區間）高出切截標準
- 長期教育下，有一定的生活自理能力
- 具溝通互動能力

❖ 此時，適應行為的多元評量很重要！！

小結

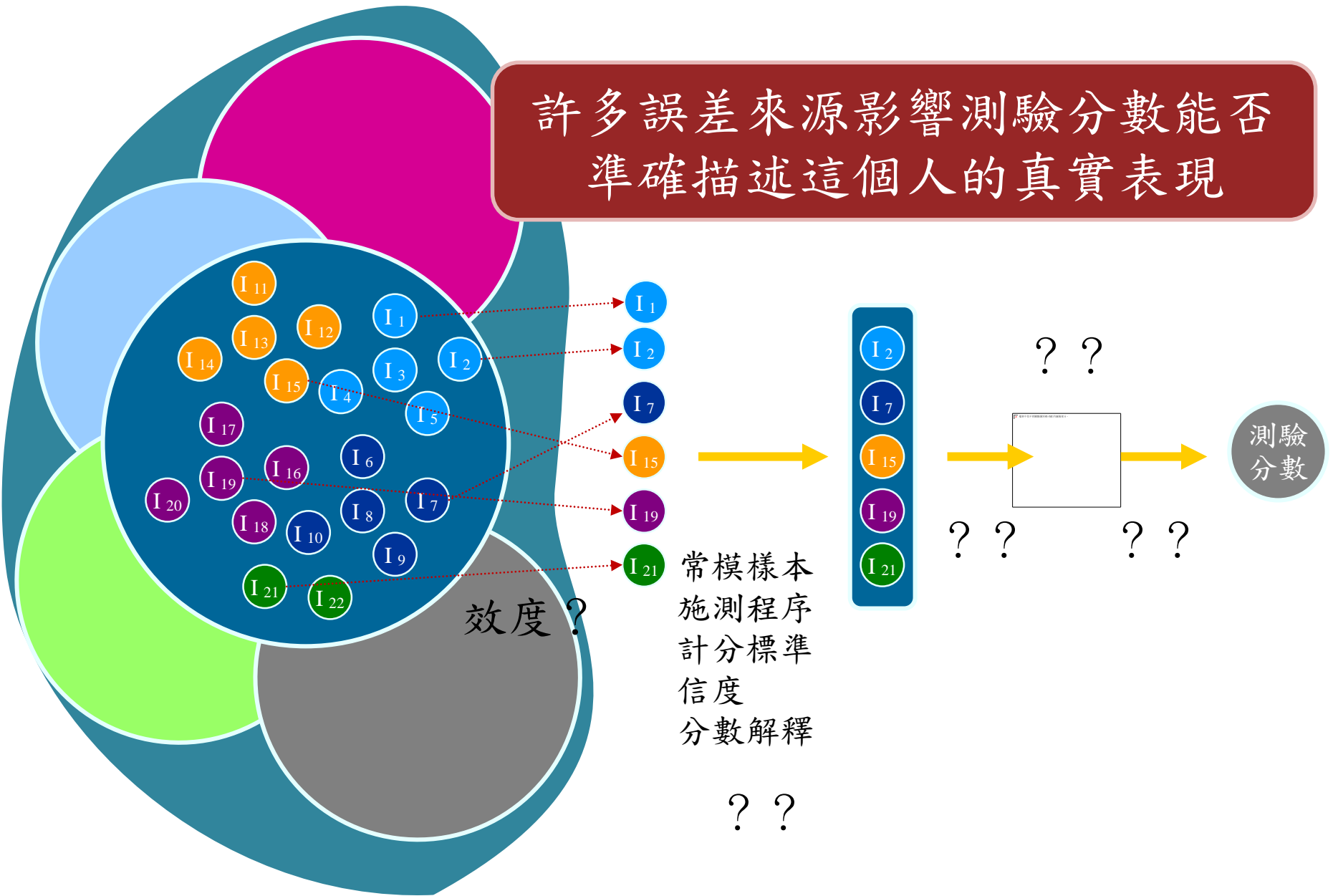
❖ 以上簡報討論哪些議題

類推應用

❖ 要研判為學習障礙的識字困難亞型

- 需要測量哪些構念
- 可以使用哪些測驗？
 - ▶ 測量什麼？有何限制？
- 還需要哪些資料？
- 如何說明符合顯著困難的標準？

許多誤差來源影響測驗分數能否
準確描述這個人的真實表現



標準化測驗編製的歷程

測驗的選擇

❖ 鑑定時，常常需要施測標準化測驗，以得到量化數據做為研判是否達顯著困難的參考，但選擇測驗要留意：

- 選擇「對」（適合）的測驗：
 - ▶ 測驗的內容是否合適：與要測量的構念相近
 - ▶ 測驗是否適合受測的學生？如施測/反應方式的限制，以致無法測得學生真實表現。
- 儘量選擇「好」的測驗：
 - ▶ 能夠讓你量得準（信度）：測驗分數的一致性 or 穩定性。
 - ▶ 測得對（效度）：測量欲測量的構念。
 - ▶ 分數對得準（常模）：常模適切。

除了選對使用標準化測驗時，還要：

- ❖ 正確施測
- ❖ 適當的解釋測驗分數：
 - 正確的說明與標準化常模比較後個案的表現。
- ❖ 施測時觀察反應歷程、錯誤類型也很重要。
- ❖ 最後，除了測驗分數，還要參考其他評量資料做綜合研判。

國內可用測驗的品質不一，
解釋時更要謹慎！

正確施測

❖ 一定要詳細閱讀測驗指導手冊後施測

- 標準化：正確的施測方式、實施程序、計時、計分
- 如因個案因素需中止、分段實施、調整順序，應記錄。
- 如要調整評量方式（如：自己讀改報讀、限時改不限時完成等）以取得更多個案反應資料，應在標準化程序後再實施。

正確施測：以適應行為的評量為例

❖ 適應行為、情緒、行為等由他人檢核的量表

- 評定「實際上會不會做」，不是「能做卻不去做」
- 應由熟悉個案且長期接觸的人評定，如心評本人、教師或家長。
- 心評不是計算分數的工具，對他人評定結果必須驗證，包括和自己經由觀察、晤談、教學所認識的個案驗證。
- 評定好不好、適不適當的參照的標準是實足年齡階段一般學生的行為。
- 這些檢核表、量表能幫助我們瞭解這個構念；能指出行為，但從行為徵候做診斷需訓練。

正確施測：以適應行為的評量為例

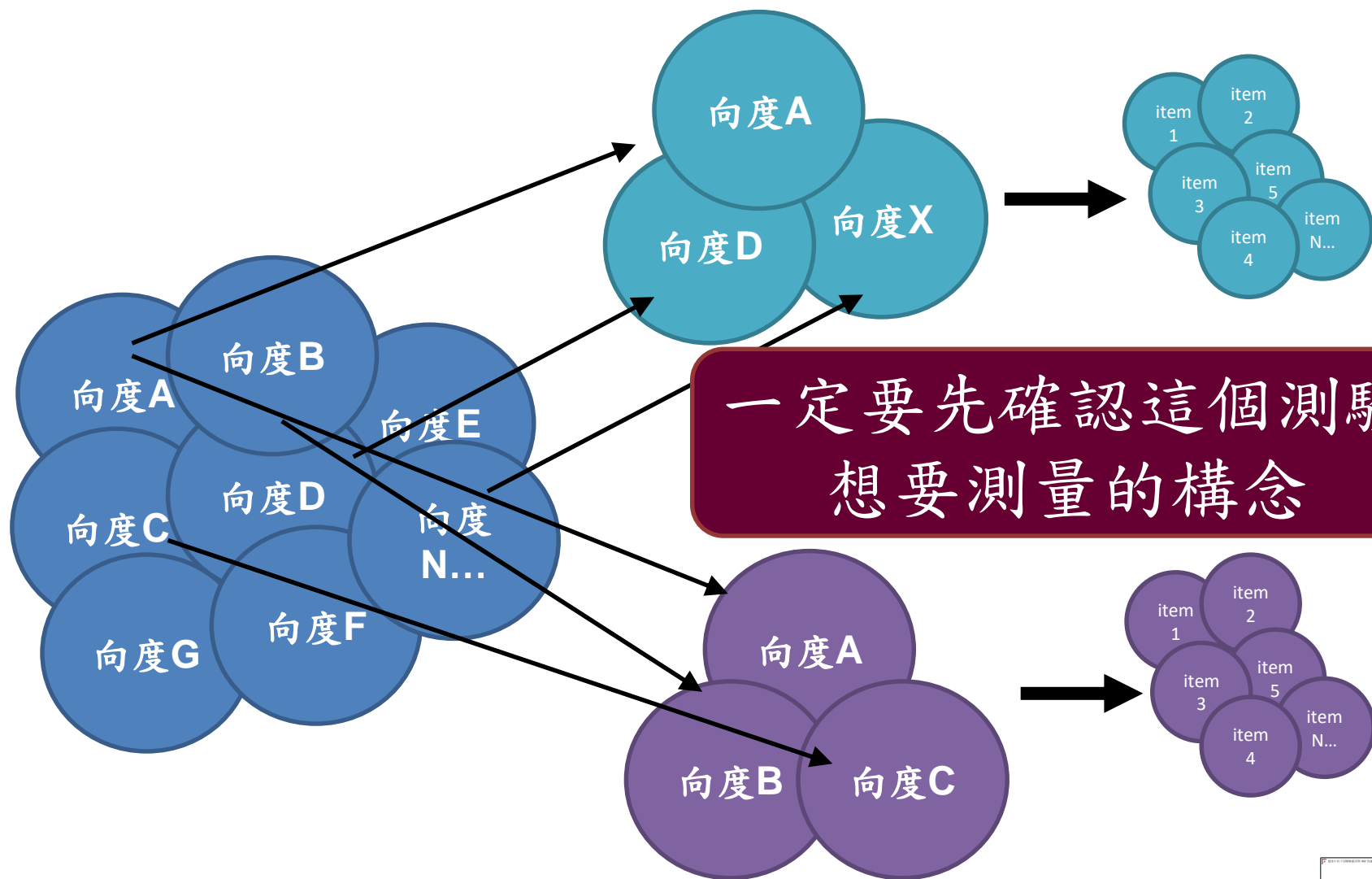
❖ 仔細參照指導手冊 說明給分

計分標準 \ 項目	3. 會清洗食物、餐具 (如：水果、蔬菜、碗盤、鍋子等)。
4分	能自行完成至少四種食物、餐具的清洗工作，且清洗得很乾淨。
3分	①在偶而協助下能完成至少四種食物、餐具的清洗工作。 ②能自行完成三種食物、餐具的清洗工作，且清洗得很乾淨。
2分	①在約一半過程的協助下能完成至少四種食物、餐具的清洗工作。 ②能自行完成兩種食物、餐具的清洗工作，且清洗得很乾淨。
1分	①在大部分的協助下能完成至少四種食物、餐具的清洗工作。 ②能自行完成一種食物、餐具的清洗工作，且清洗得很乾淨。
0分	完全不會清洗食物、餐具。

計分標準 \ 項目	4. 會使用廚房用具 (如：冰箱、瓦斯爐、菜刀、微波爐等)。
4分	能自行正確使用至少四種廚房用具。
3分	①在偶而協助下能使用至少四種廚房用具。 ②能自行正確使用三種廚房用具。
2分	①在約一半過程的協助下能使用至少四種廚房用具。 ②能自行正確使用兩種廚房用具。
1分	①在大部分的協助下能使用至少四種廚房用具。 ②能自行正確使用一種廚房用具。
0分	完全不會使用廚房用具。

計分標準 \ 項目	28. 會處理青春期時所產生的生理狀況 (未到青春期者請勾選0分)。
4分	男生：能自行 a.清洗生殖器官 b.刮鬍子 c.處理青春痘 d.處理夢遺 女生：能自行 a.清洗生殖器官 b.穿胸罩 c.處理青春痘 d.處理月經
3分	①在偶而協助下能做好依其性別所列的四項處理工作。 ②能自行做好依其性別所列四項處理工作中的三項。
2分	①在約一半過程的協助下能做好依其性別所列的四項處理工作。 ②能自行做好依其性別所列四項處理工作中的兩項。
1分	①在大部分的協助下能做好依其性別所列的四項處理工作。 ②能自行做好依其性別所列四項處理工作中的一項。
0分	完全不會處理青春期時所產生的生理狀況。

心理計量基本概念—效度



效度—首先一定要知道這個測驗在測量什麼

❖ 中華「適應行為」量表：

- 居家生活：生活自理、家事技能；學校生活：溝通能力、實用知識、獨立自主、安全衛生；社區生活：社區活動、消費技能、社會技能、休閒活動；工作：動作發展狀況、工作活動、社會-工作行為

❖ 「學校適應」能力量表：

- 感官能力：視力、聽力；生理動作能力：身體狀況、自我照顧、基本行動、動作協調；社會情緒能力：情緒、環境適應、團體生活；語言能力：語言理解、語言表達；學習能力及成就：學習能力、學習態度與習慣、學科成就

❖ 「社會適應」表現檢核表：

- 自我照顧、動作、溝通、社會情緒、學科學習

向度不完全同
評量目的也不盡相同

❖ 文蘭「適應行為」量表：

- 溝通：接受性、表達性、讀寫；日常生活技巧：個人、家庭、社區；社會化：人際關係、遊戲與休閒、應對進退技巧；動作技巧：粗大動作、精細動作

❖ 「適應行為」\評量系統第二版（ABAS-II）

- 溝通、學習功能、自我引導（概念知能）；休閒、社交（社會知能）；社區應用、學校生活、健康與安全、自我照顧（實用技巧）

心理計量基本概念—效度

測驗名稱	反應方式	構念（識字）
<u>中文年級認字量表</u> （黃秀霜，2001）	<ul style="list-style-type: none"> • 讀出該字 或 • 寫出注音 	<ul style="list-style-type: none"> • 表現 <ul style="list-style-type: none"> • 識字正確性 • <u>識字流暢性</u> • 書寫（國字、注音）？ • 認知層面 <ul style="list-style-type: none"> • 解碼 • 字詞辨識 • 注音（聲韻覺識） • 詞彙提取 • 部件及組字規則？
<u>識字量評估測驗</u> （洪儷瑜等，2007）	<ul style="list-style-type: none"> • 看國字寫出注音 • 看國字造詞（書寫） 	
<u>自編識字量估計測驗</u> （李俊仁）	<ul style="list-style-type: none"> • 看字讀音 • 看字造詞（口頭） 	
<u>常見字流暢性測驗</u> （洪儷瑜等，2007）	<ul style="list-style-type: none"> • 看字讀音 • 看字造詞（口頭） 	

心理計量基本概念—效度

❖ 能針對設定的測量構念有效的評估

- 如，以切截分數（cutoff score）找出閱讀障礙，要呈現此分數是有效的證據。
- 如，測驗宣稱能幫助課程規劃，則其測驗分數跟剖面圖能準（item/task）確找到教學起始點。

❖ 構念—向度—項目（item/task）代表性

❖ 受試在項目的反應歷程可代表構念本身反應歷程？

❖ 構念的結構與比重

❖ 與其他變因（效標）的相關性

❖ 整個評量過程都涉及效度：

- 測驗編製、實施情況、學生表現歷程、評分所得結果、結果詮釋、詮釋產生的後果、整個過程傳遞的價值

心理計量基本概念——信度種類

❖ 信度是指測驗所得到結果的一致性或穩定性，而非測驗或量表本身。沒信度沒效度。

● 穩定性-

▶ 重測信度

- 影響因素：個體實質變化、再測練習效果、評量特質短時間內快速變化（心情、情緒）

● 可重複性-

▶ 複本信度

- 估計的誤差：時間誤差、內容取樣誤差

▶ 內部一致性係數：折半法、庫李信度、 α 係數

- 不能估計速度測驗
- 估計的誤差：內容取樣、內容異質性

● 觀察者間一致性

- ▶ 由他人評分的量表

知道信度概念要做什麼？

心理計量基本概念——信度評鑑

❖ 評量目的：

- 影響大，無法再測，信度要高
- 不同評分者挑出前10名，重點在相對位置的給分標準不變，給分寬鬆不影響；決定畢業與否，給分寬鬆影響大，評分者給分一致性重要
- 預測，穩定係數重要
- 學習成果：內容取樣誤差重要

❖ 影響信度因素：

- 評量項目多寡、團體異質性、信度估計方法

心理計量基本概念—信度評鑑

信度係數	使用注意事項
0.60-0.65	最好不要用
0.65-0.70	最低可接受值
0.70-0.80	不錯
0.80-0.90	非常好

分量表內不一致性低於0.60，總量表低於0.80建議重編

DeVellis (1991)

不要被統計顯著性**騙了，**N**夠大，係數小也可以達到統計顯著性，但決定係數 (r^2) 小，可解釋量不高

心理計量基本概念——常模類型與適切性

❖ 常模類型

- 全國性常模
- 地區性常模
- 特殊團體常模
- 學校平均數常模

- 使用測驗時，你會檢視常模樣本數、代表性等資料？
- 如果常模有問題，分數還有多少準確性？

國內可用測驗的品質不一，
解釋時更要謹慎！

❖ 常模的適切性

- 代表性與樣本大小
- 新近性
- 適切性：測驗對象與常模團體性質接近

❖ 好的常模才能讓測驗分數對得準

解釋測驗的表現—常模參照解釋

❖ 發展分數

- 年齡當量
- 年級當量

❖ 相對地位分數

- 百分位分數
- 標準分數

解釋測驗的表現

❖ 標準參照解釋 (Standards-referenced)

- 國家課程標準 (領綱分年學習素養)
- 以多元評量方式，專業判斷學生是否進步/達到分年預期的參照點 (學習表現)。

❖ 效標參照解釋 (Criterion-referenced)

- 判斷學生在特定領域 (domain) 的表現是否達到預設的校標 (benchmarks)。
- 明確的內容範圍及能具體表現該內容範圍的行為校標。

❖ 常模參照解釋 (Norm-referenced)

- 決定學生在群體中的相對位置。

解釋測驗的表現—常模參照解釋

❖ 發展（成長）分數：

- 年齡分數（年齡當量）：
 - ▶ 將常模樣本分成各年齡組，每一年齡組原始分數的平均數即年齡常模。
 - ▶ 在A測驗上，普通9歲兒童的原始分數（如12分）相當於九歲組平均數（12）時，可轉換成年齡當量9。
 - ▶ 一個6歲或11歲兒童在同一測驗得到相同的原始分數（如12分），年齡當量也是9。
- 年級分數（年級當量）相似

哪些測驗使用年級當量？

解釋測驗的表現—常模參照解釋

- 發展分數使用注意事項：
 - ▶ 適用隨年齡/年級逐漸增加的特質
 - ▶ 年齡/年級當量的單位不相等； $4.0 \rightarrow 5.0 \neq 7.0 \rightarrow 8.0$
 - ▶ 所得年齡/年級當量不等於有該年齡/年級的相同能力
- 三年級學生在數學成就測驗的年級當量為 6，這表示：
 - ▶ 該兒童具有六年級的數學程度。(X)
 - ▶ 該兒童在此測驗答對的題數和常模中六年級學生答對的平均題數相同。
 - ▶ 學生和同年級相較，數學表現較優越，但不知道多優越！

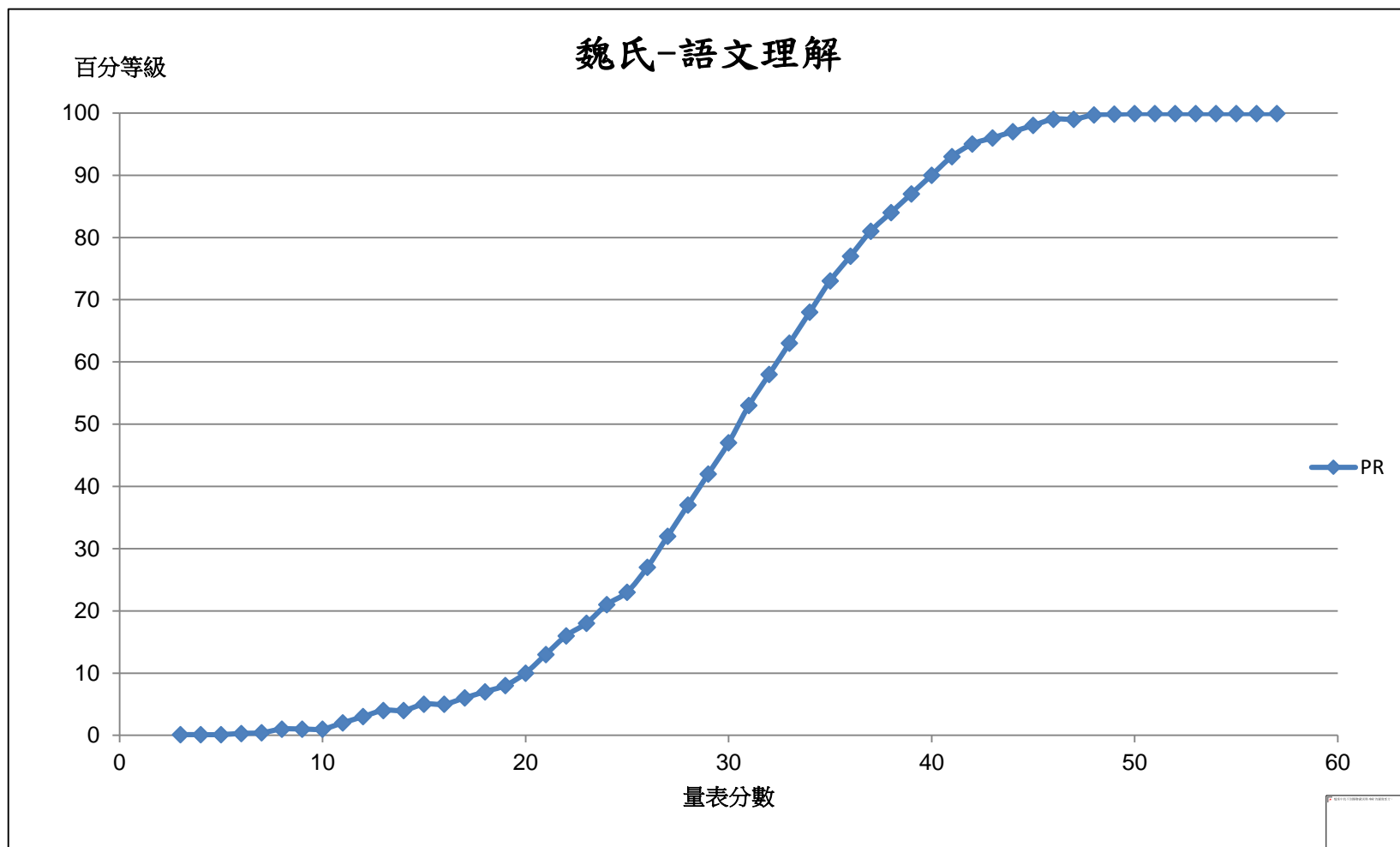
解釋測驗的表現—常模參照解釋

❖ 百分位分數

- 比較受試者得分在測驗上的「等第」。
- 在一常模樣本中，位於某一原始分數以下人數的百分比
- 百分位數與百分等級：以測驗分數界定百分點時稱百分位數；以百分點指明該點以下包含人數的百分率時，稱為百分等級。
 - ◆ 例如：第54百分位數的原始分數為130，或原始分數130的百分等級為54（PR54）
- 百分等級和答對測驗問題的%不同。 $25/50 = 50\% \neq \text{PR}50$
- 百分等級是順序量尺，不是等距量尺；分配的中央和兩端，原始分數的差相同，百分等級差很多。

解釋測驗的表現—常模參照解釋

❖ 常態分配下百分等級與量表分數關係



解釋測驗的表現—常模參照解釋

❖ 標準分數家族

- 以標準差為單位，表示個體分數到平均數間距離的一種轉換分數

- z 分數 = $(X-M) / SD$ ， $SD = \sqrt{\frac{\sum(X-M)^2}{N-1}}$

- T 分數 = $10z+50$

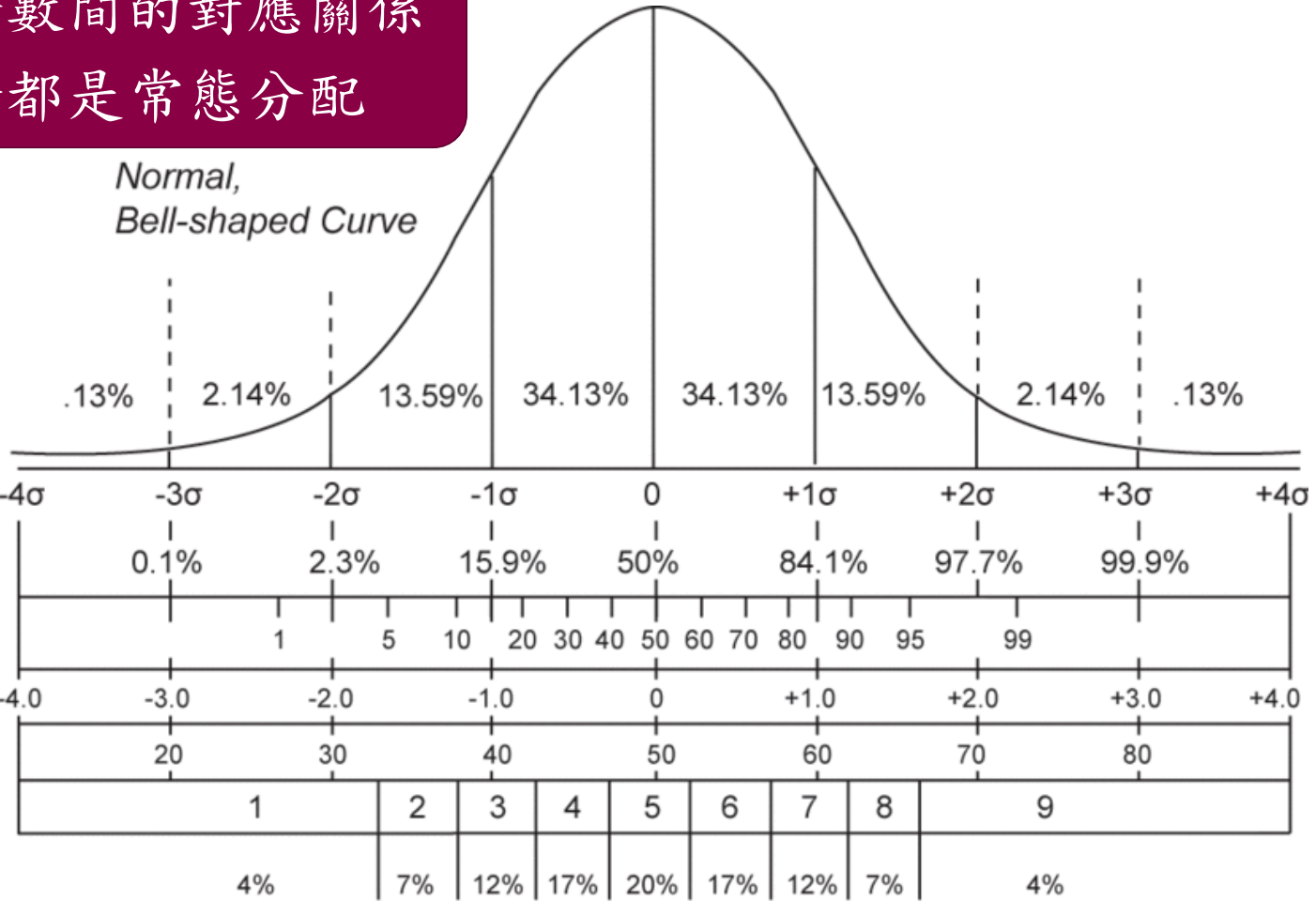
- 離差智商，如WISC：平均數100，標準差15

常態分配下不同相對地位分數間的關係

- 請記熟不同分數間的對應關係
- 不是每種構念都是常態分配

Normal,
Bell-shaped Curve

常態分配曲線人數百分比
Percentage of cases in 8 portions of the curve



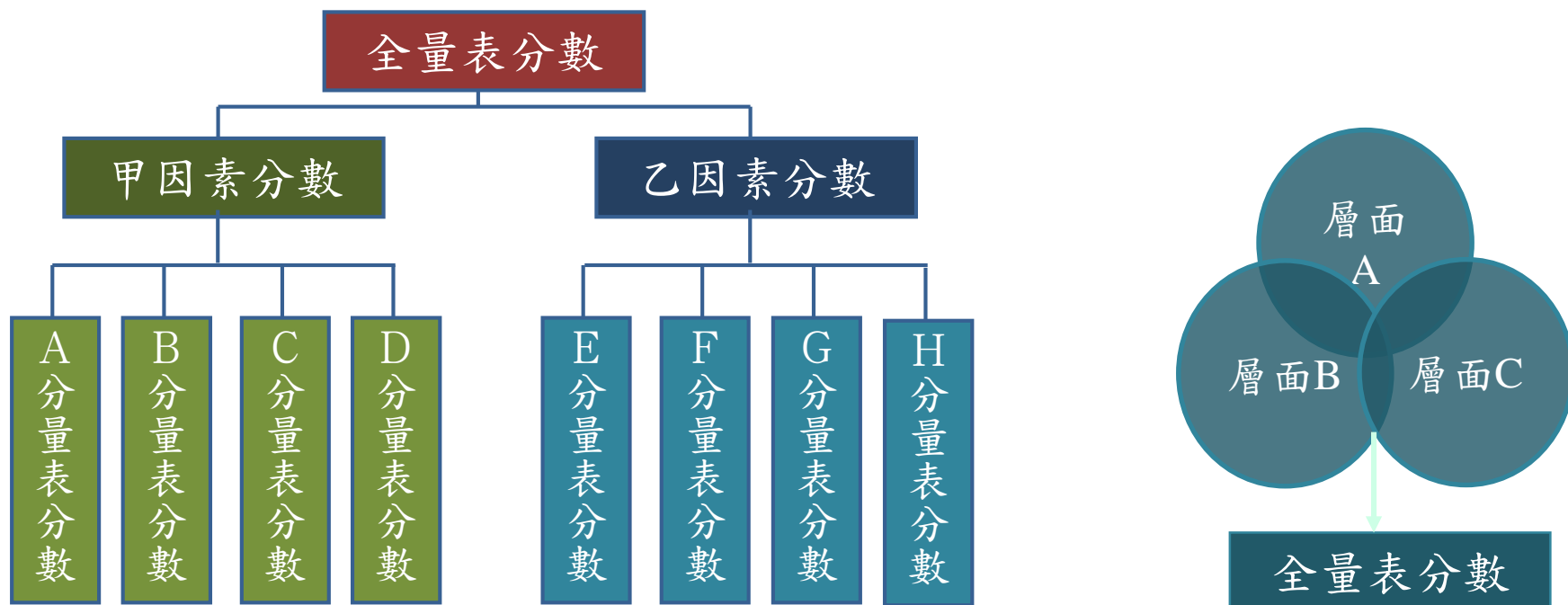
解釋測驗的表現—常模參照解釋

❖ 心評基本功

Z分數	-3	-2	-1	0	1	2	3
T分數	20	30	40	50	60	70	80
PR	<1	2	16	50	84	98	99+
IQ 指數	55	70	85	100	115	130	145
量表 分數	1	4	7	10	13	16	19

- 同一構念，不同測驗分數相互參照一致性（無法直接比較）
- **分數是常態分配下**，不同能力之測驗表現的相互參照

解釋測驗的表現—常模參照解釋



❖ 只能解釋全量表分數還是能分別解釋個別因素（分量表）分數/指數？

❖ 如果可以分別解釋，什麼時機用全量表或因素？

解釋測驗的表現—多嚴重才符合障礙？

❖ 顯著困難常用參照方式（和一般學生比較）

聽覺理解、口語表達、識字、閱讀理解、書寫、數學運算、注意力、記憶、動作協調

- 年級差距：如5年級生但低於3年級的平均數。
- 標準分數差距：如低於平均數負X個標準差
- 百分等級常模：如PR15、PR3以下。
- 切截分數：表現低於建議之切截分數。
 - ▶ 注意測驗訂切截分數採用的標準是多少？很低嗎？通常要避免錯誤拒絕，所以...
- 應付該學習所需的基本量：學障基本技能常用到

❖ 對到正確的常模：常模年級/年齡範圍、代表性

❖ 不是符合此基本參考標準就符合障礙資格，仍須參考其他資料綜合研判。

解釋測驗的表現—多嚴重才符合障礙？

- ❖ 有些構念，適合用標準分數差距、年級差距來判斷；有些構念不是適合這些方法。
 - 智力、閱讀理解、學科成就等，適合（Why？）。
 - 基本技能：識字、書寫、基本計算...等，不一定適合。
- ❖ 對到正確的常模：常模年級/年齡範圍、代表性。
- ❖ 不是達到此基本參考標準就符合障礙資格，仍須參考其他資料綜合研判。

解釋測驗的表現—多嚴重才符合障礙？

❖ 智力正常

- 以何種方式決定正常？
 - ▶ 標準化個別智力測驗、標準化測驗、日常表現...？
- 如以測驗決定，智力正常的值要訂在多少？
 - ▶ 70（不是智障就是正常）？75？80？85（-1SD）
 - ▶ 不同值影響排除的範圍
 - ▶ 不同亞型影響「值」：在學障，基本技能可採比較低的標準，和理解有關的要較高點的智力標準

❖ 哪些障礙類別有智力正常的標準？一樣嗎？

解釋測驗的表現—多嚴重才符合障礙？

❖ 智力正常

- 對於這個值，採點估計還是區間估計？
 - ▶ 採信賴區間，信賴水準採.01/.05影響區間分數帶大小
- 哪些類型的測驗可以決定？
 - ▶ **WISC4 ? TONI ? PPVT ?**
 - ▶ 接受魏氏因素指數？全部？哪個？
- 「測驗分數」能否真正估計學生「真實」智力
 - ▶ 施測程序沒錯、分數沒算錯、常模沒對錯
 - ▶ **是否低估真實潛能？尤其是高年級以後**
 - 智力測驗與學校表現（學習累積）高相關，是與同儕比較，以目前表現預測未來成就；學障問題影響學習累積。

多元評量

多嚴重才符合障礙？學障鑑定基準補充說明

❖ 顯著困難（聽、說、識字、寫、算等基本技能）：

- 須留意該測驗之編製特性、測驗使用限制、測量誤差來源及欲測量能力或表現的發展特性。

什麼意思？
- 參照日常生活與學習表現，勿單以測驗結果拒絕或接受研判學習障礙。
 - ▶ 標準分數表現低於其年級（級）組平均數達2個標準差以上或在該年級常模百分等級3以下者。
 - ▶ 年級/年齡分數得分低於就讀年級1個年級/1歲之平均數（年級分數/當量）以上，且不同年段差距不同：
 - 低年級：落後1個年級或1歲以上。
 - 中年級：1.5個年級或1.5歲以上。
 - 高年級：2個年級或2歲以上。
 - 國中：未達國小四年級平均數。

多嚴重才符合障礙？學障鑑定基準補充說明

各年級估計字量之標準誤及 90% 和 95% 信賴區間

年段	估計標準誤	平均數	標準差	90%信賴區間	95%信賴區間
小一	174	712.37	444.99	426~998	371~1,053
小二	142	1,248.57	363.54	1,015~1,483	971~1,527
小三	356	2,108.04	816.30	1,522~2,694	1,410~2,806
小四	450	2,660.52	855.00	1,921~3,401	1,779~3,543
小五	366	3,142.08	996.16	2,540~3,744	2,425~3,859
小六	413	3,340.02	998.47	2,661~4,019	2,531~4,149
國一	414	3,547.97	998.11	2,867~4,229	2,737~4,359
國二	548	3,521.06	1,039.71	2,620~4,422	2,447~4,595
國三	449	3,747.34	1,081.88	3,008~4,486	2,867~4,627

資料來源：王瓊珠等，2008：一到九年級學生國字識字量發展

失學國民脫盲識字標準：**1680字**；一般成人日常生活所需之基本字彙為**2,328字**（黃富順，1994）

在現代，標準該提高還是下降？

多嚴重才符合障礙？學障鑑定基準補充說明

❖ 顯著困難（聽、說、讀、寫、算等技能）：

- 對高年級以上學生，使用基本技能相關測驗未達上述顯著困難標準（如PR3至PR25）不表示學生沒有困難，須透過其他資料佐證並與學習歷程、學業相關表現比較後綜合研判。

Why?

- 有些篩選性測驗會建議達篩選標準的切截分數。使用時要注意，編製者是否為了要減少錯誤拒絕而建議較寬鬆的標準，如設PR25、PR15的對照分數做為切截標準。因此，學生得分在切截分數以下，不一定就有顯著困難，仍要參考其他學習表現。

多嚴重才符合障礙？學障鑑定基準補充說明

❖ 能力差距的使用時機（內在能力的比較）：

- 如差距分數**30**分以上是標準，這些標準分數：**A**生智力**135**，閱讀**95**；**B**生智力**95**，閱讀**70**，誰是閱讀障礙？
- 是間接推論，不是直接證據
- 兩高相關能力間的表現不一致才有意義，如智力和閱讀理解（但兩構念越相像，差距分數的信度越差，迴歸）
- 差距達到統計顯著性，不代表真有問題，如差距**10**分雖達統計顯著，但可能有**25%**的人有這種情形
- 智力測驗分測驗間/學科間表現顯著差距，不宜做判斷標準。人本來就有能力間的差距（如果支持多元智能）
- 所以，它是參考的標準，需參考其他資料綜合研判



❖ 分數間差異的統計顯著性 VS. 母群體差異的基本率（發生頻率）的意涵

- 差異是否真實而不是測量誤差？
- 差異雖然真實，但是否具臨床的意義？

表乙.1 各年齡組及全體樣本所得指數分數間差異值達.15 和.05 統計顯著水準所要求的臨界值 (台灣樣本)

年齡組	顯著水準	指數配對					
		語文理解 vs. 知覺推理	語文理解 vs. 工作記憶	語文理解 vs. 處理速度	知覺推理 vs. 工作記憶	知覺推理 vs. 處理速度	工作記憶 vs. 處理速度
6:0-6:11	.15	8.99	9.31	11.59	7.84	10.45	10.73
	.05	12.23	12.67	15.78	10.67	14.22	14.60
7:0-7:11	.15	9.35	8.84	11.21	8.20	10.71	10.28
	.05	12.72	12.03	15.26	11.16	14.58	13.99
8:0-8:11	.15	8.24	8.33	10.28	8.51	10.43	10.50
	.05	11.22	11.34	13.99	11.58	14.19	14.29
9:0-9:11	.15	8.08	7.29	10.32	7.63	10.56	9.97
	.05	10.99	9.93	14.05	10.38	14.37	13.57
10:0-10:11	.15	8.73	7.33	10.27	8.71	11.29	10.25
	.05	11.89	9.98	13.98	11.85	15.37	13.95
11:0-11:11	.15	9.95	8.09	10.38	10.11	12.02	10.54
	.05	13.54	11.02	14.13	13.76	16.36	14.34
12:0-12:11	.15	7.96	8.02	8.80	9.25	9.94	9.99
	.05	10.83	10.91	11.97	12.60	13.53	13.59
13:0-13:11	.15	8.49	7.61	9.19	8.70	10.11	9.38
	.05	11.55	10.35	12.51	11.84	13.76	12.77
14:0-14:11	.15	8.11	8.45	9.47	8.93	9.90	10.18
	.05	11.04	11.50	12.89	12.15	13.48	13.85
15:0-15:11	.15	8.97	7.90	9.00	9.73	10.64	9.75
	.05	12.21	10.75	12.25	13.24	14.48	13.28
16:0-16:11	.15	8.84	7.01	9.17	9.08	10.83	9.40
	.05	12.03	9.55	12.48	12.36	14.74	12.79
所有年齡組	.15	8.72	8.04	10.01	8.82	10.64	10.10
	.05	11.87	10.95	13.63	12.00	14.48	13.74

註：達統計顯著水準所要求的差異是依據各年齡組的各指數之測量標準誤計算而得。其計算公式為：
 差異分數的臨界值 = $Z \sqrt{SEM_a^2 + SEM_b^2}$ 。其中 Z 為不同顯著水準下所對應的常態分配標準分數。
 SEM_a 和 SEM_b 分別為兩指數之測量標準誤。

表乙.2-1 全部標準化樣本所得六對指數分數間差異值的累積百分比 (%)

差異值	VCI vs. PRI		VCI vs. WMI		VCI vs. PSI		PRI vs. WMI	
	VCI < PRI (-)	VCI > PRI (+)	VCI < WMI (-)	VCI > WMI (+)	VCI < PSI (-)	VCI > PSI (+)	PRI < WMI (-)	PRI > WMI (+)
≥40	0.2	0.6	0.3	0.1	0.6	1.1	0.2	0.2
39	0.2	0.6	0.6	0.1	0.7	1.5	0.2	0.3
38	0.2	0.7	0.7	0.5	0.9	1.5	0.2	0.3
37	0.3	0.7	0.9	0.6	0.9	1.5	0.2	0.6
36	0.4	0.7	1.1	0.6	0.9	2.0	0.2	0.6
35	0.8	0.7	1.1	0.9	1.2	2.4	0.2	0.9
34	0.9	0.8	1.2	1.0	1.3	2.6	0.3	1.1
33	1.4	0.9	1.2	1.3	1.9	3.5	0.4	1.2
32	1.9	1.2	1.3	1.7	2.4	3.6	0.6	1.5
31	2.9	1.4	1.4	2.2	2.6	4.0	0.7	1.8
30	2.2	2.0	2.0	2.5	3.2	4.6	0.9	1.9
29	2.6	2.1	2.0	2.8	3.5	4.8	1.3	2.3
28	3.6	2.7	2.3	3.5	4.1	5.3	1.7	2.5
27	3.8	2.8	3.1	3.6	4.4	5.8	2.0	2.6
26	4.2	3.6	3.6	3.8	5.6	6.3	2.5	3.1
25	4.6	3.9	4.1	4.4	6.2	7.4	3.4	3.7
24	5.6	4.9	4.5	5.1	6.7	8.5	4.0	4.3
23	6.3	5.1	5.4	5.7	7.6	9.1	5.0	5.1
22	7.1	5.9	6.3	6.7	8.8	10.0	6.3	5.7
21	7.6	6.2	6.9	7.1	10.0	11.1	7.0	7.1
20	8.6	8.0	8.1	8.3	11.6	11.9	7.9	8.1
19	9.6	8.9	9.4	9.5	12.4	13.3	9.4	9.0
18	10.0	9.7	10.3	10.3	13.9	14.5	9.8	10.1
17	11.4	10.1	11.6	11.4	15.9	15.8	12.3	11.4
16	13.4	11.9	13.4	13.9	18.2	17.0	13.4	12.6
15	15.4	12.6	16.0	15.8	18.7	18.5	15.4	13.7
14	18.5	17.1	17.5	17.9	20.6	20.5	17.3	15.1
13	19.3	17.5	19.1	19.4	22.0	21.1	18.5	17.0
12	21.7	20.9	21.3	21.1	23.5	22.5	19.8	19.4
11	22.6	21.6	23.2	23.8	25.3	25.4	22.2	21.2
10	24.5	25.5	25.1	26.1	27.4	27.1	23.5	24.5
9	26.3	27.3	27.8	27.7	29.3	28.9	26.3	26.4
8	28.8	30.5	30.3	30.2	32.6	30.7	29.8	28.8
7	30.4	31.4	33.3	32.5	34.4	31.9	31.5	32.1
6	35.6	34.6	36.4	35.0	37.5	34.2	35.8	35.2
5	37.2	38.4	38.9	37.2	39.5	36.8	37.2	38.5
4	41.0	40.3	41.3	38.5	42.6	38.5	39.3	40.3
3	42.5	41.3	45.1	40.6	45.4	41.2	44.1	42.3
2	45.5	46.8	48.0	43.6	47.6	43.5	45.4	45.8
1	47.2	48.3	51.4	46.0	50.6	46.6	48.8	47.6
平均數	12.0	11.4	11.3	12.3	12.8	13.8	11.2	11.5
標準差	8.6	8.4	8.5	8.6	9.7	10.5	8.0	8.3
中位數	10.0	10.0	9.0	11.0	10.5	11.0	9.0	10.0

一些提醒

❖ 評量資料間不一致（人常是一致的，但不包括個體內在差異）：

- 同一測驗不同施測時間點
- 同類型測驗
- 同向度，不同資料來源之間

❖ 做判斷時，請先檢視：

- 測驗真的測量想要測的構念？
- 有無施測錯誤？
- 分數算錯？
- 施測時，學生的表現典型嗎？有干擾因素？

一些提醒

❖ 請小心

- 請檢查測驗的品質，尤其是效度、常模及解釋注意事項
- 對單一危險因子的過度推論
- 誤認偶發事件為有代表性的行為樣本
- 亂槍打鳥或套裝式收集資料
- 解釋行為表現的參照點錯誤

一些提醒

❖ 請記得您的評量（或施測該測驗）目的：

- 得到一個分數看符不符合鑑定標準？
- 瞭解孩子在某能力向度的表現水準？
- 分析主要困難問題以引導教學？

❖ 這個測驗能幫助您回答以上的問題？

測驗選擇

還是要回到效度的核心概念