

關於標準分數

新制「日本語能力試験」(新制測驗)為了更公平的評量考生的日語能力和測驗結果(得分)而採用「尺度得點」(標準分數)來計算得分。利用標準分數,無論考生參加哪一次測驗都能在同一標準下測量其日語能力。標準分數的優點及其計算過程之說明如下:

【標準分數的優點】

在出題過程中,即使經過專家的嚴密分析與研究,也很難保證每次試題的難易度會完全一致。是故,如果單純依答對題數計算考生所得之「粗分」來表示得分的話,測驗得分便會因為每次測驗難易度之不同而有所差異。如此一來,受測的時間點不同或者試題的難易度不同時,即使是能力相同的考生,所得的分數也不會相同。因此,我們必須考慮到用粗分表示考生日語能力時,有可能受到試題難易度的影響,以致無法如實反映出考生實際日語能力的問題。

相較之下,新制測驗採用的「尺度得點」(標準分數)在測量能力上有其優點,即「日語能力評量結果不受試題難易度的影響,且可依據共同量尺將之數值化。」因此,日語能力相同的考生,不論參加哪一次測驗,得分可能都相同。採用「尺度得點」(標準分數)後,只要是相同級數、相同測驗項目,兩次不同測驗中所算出的「標準分數的差異」可不受測驗難易度之影響而可解釋為「日語能力的差異」。

【標準分數的計算過程】

這種不會因測驗難易度影響的「尺度得點」(標準分數)之計算方式,是以統計學理論「試題反映理論」(Item Response Theory: IRT)來運算。此計算方式與計算答對題數所得出之分數(粗分)完全不同。

「尺度得點」(標準分數)是由各題之答題狀況(答對或答錯)、也就是由「答題組型」得出數理性結果。例如:10個題目(項目)組成的測驗中,最多有1024組(2^{10} 組)答題組型。在新制的「尺度得點」(標準分數)計算過程中,根據這些答題組型,在每一得分範圍0~60分的刻度量尺上衡量考生的日語能力(N4、N5的「言語知識<文字・語彙・文法>、讀解」則為0~120分)。由於把10個題目總共多達1024組的答題組型(實際測驗會更多)分成了61組,不同的答題組型有時會得到相同的標準分數。因此,如有兩位考生其答對題數及答題組型雖不同,其「尺度得點」(標準分數)可能相同。反之,即使答對題數相同,其「尺度得點」(標準分數)有時也會因答題組型不同而有所差異。

關於「尺度得點」(標準分數)計算過程的詳細情況,請參照下列文獻。

【文獻】

池田央 (1994) 『現代テスト理論』 朝倉書店

豊田秀樹 (2002) 『項目反応理論[入門編] -テストと測定の科学-』 朝倉書店

渡辺直登・野口裕之 (編) (1999) 『組織心理測定論-項目反応理論のフロンティア-』 白桃書房

關於合格基準的設定

新制的合格基準，是依照統計分析的結果以及與舊制的合格基準整合後所設定的。由此，能夠通過舊制 1 級、2 級、3 級、4 級的考生，可以視其日語能力應該也能夠通過新制的 N1、N2、N4、N5。至於舊制中沒有的 N3，乃是統計分析通過舊制 2 級和 3 級の日語能力，在其兩者的通過標準之間，設定了 N3 的合格基準。