

## 今年文憑試的試題容易嗎？——誤解與等級釐定

最近文憑試正在進行，傳媒透過訪問學生及老師等，希望了解今年試題是否容易，同學更以此推論是否更易及格或獲取更佳成績。這反映公眾對等級釐定及公開考試運作存在誤解。

### 評分標準也決定生死

只看一份試卷的題目，我們難以決定是否更易及格或獲取較佳成績。就算題目看來容易，但評分要求可能很高，極難得分。

同理，當學生可以從多題中選答一題時，就算他們對所有題目都有相同信心，但選答看來較易的題目，卻不一定更為有利，因為它們的評分標準及要求可能更高。

### 「拉曲線」與標準參照

在近年教育改革中，我們常強調已經不再用「拉曲線（curve）」而改以標準參照，來定及格及各級的標準。也即是說，我們不預先設定每個成績級別的人數比例，而是用一些學業表現指標（教學目標），去決定某學生的成績級別。其實這是一個很難具體操作或實踐的教育理念，極可能只是夢想。

首先，在某份試卷獲取 90 分、80 分還是 70 分，才算達到「充分掌握」某個學習概念？可稱得上優良？這決定是誰說了算？明天睡醒，這「權威」的決定還是一樣嗎？將一些教學目標及要求，變成某一個考試的分數，有一些行業內常用方法，但絕對並不像物理實驗室內一個極為客觀的過程。

其次，教師並無能力將每年試題調校至同一難度水準，所以今年考得 80 分，與去年 80 分的能力，並不相同。在日常校內考試，我們無奈接受這些局限及誤差。在大型公開考試，我們難以容忍。就算每年相差數分，數年累積下來，誤差甚大。所以大型考試，都採用複雜統計方法，以保證跨不同年份的成績是可比的。簡單來說，如果今年試題較淺，則必須獲取較高分數才能及格。每年獲取及格（或優異）的分數線並不相同。

### 釐定等級需考慮其他因素

我們對「標準」的要求，其實也並非一成不變。我的破爛普通話，在二三十年前，尚算過關，如今與香港一般中小學生相比，常成他們的笑柄。但是公眾不希望這標準常變，故此我們用統計方法去保證這標準不變（並非保證各等級人數百分比不變），可惜公眾甚至一些教育界人士不熟悉測量理論，誤稱之為「拉曲線」並強烈反對，實為誤解。

最後，我必須強調釐定及格優良的標準，並非一個與自然科學相同的客觀過程。但某年釐定標準後，用數學方法實現跨年水準一致，卻是很多國家常用及十分科學的一個技術，令我們可以有效地監控香港整體教育水準等方面的變化，也令公開試成績在公眾中有高的認受性。

正因為標準是由人所釐定，近日聽說某學科委員說，要將他們的科目優良及格等級要求定得低一點，讓該科考生更易獲取優良的成績，那麼報讀該科的學生自然會更多。這絕對可能發生。所以除了邀請相關人士釐定標準外，我們也用統計方法，確保某一能力水準的人（對一極大群的學生而言），他們在不同學科的不同難度考試中，如果發揮正常，那麼我們就會看到他們各科的成績等級（用統計方法調整後），是接近或相當。因為上述原因與其他教育計劃及政策的需要，由相關人士所釐定的標準是一個有重要參考價值、而非唯一的考慮因素。

近年我亦帶領及參與國家頗多學業水準釐定的工作。但在香港來說，因 334 改革的相關公開考試標準釐定工作，可能是我一生中，唯一能目睹的大事，希望本文能對公眾及業界，澄清一些觀念及誤解，令有關當局能充分全盤考慮所有相關因素，去釐定這可能會延用數十年的標準。

作者是中文大學副校長