

香港中學文憑考試評級程序與水平參照成績匯報
2018 年 3 月

目錄

1. 簡介
2. 成績匯報制度
3. 等級描述
4. 評級程序
5. 設定水平
6. 維持水平
7. 匯報成績

附錄 1: 組別能力指數

附錄 2: 百分位等值法

附錄 3: 潛在特性模型

1. 簡介

香港中學文憑考試（文憑試）採用水平參照模式匯報評核結果，各個科目也會參照各自一套明確而固定的標準，來匯報考生表現所屬的等級。

香港的公開考試並不是首次採用水平參照匯報成績，由 2007 年開始，香港中學會考（會考）中國語文和英國語文科，已採用水平參照模式匯報成績，以確保評級所參照的標準獲國際認可、透明度高、明確而穩定。

為了確保年與年之間的水平穩定，並維持同一年內各科成績的可比性，香港考試及評核局（考评局）確立了一套與水平參照模式相配合的評級制度。

本小冊子旨在解釋評級程序與水平參照成績匯報的原理、方法及重點。

2 成績匯報制度

文憑試的科目可分為三類：甲類、乙類及丙類科目。

甲類科目

甲類高中科目的評級由考評局全權負責。在香港中學文憑考試證書上，考生的表現以五個等級（1—5 級）匯報，第 5 級為最高等級。在第 5 級的考生中，成績最優異的可獲評為 5** 級；成績次佳的獲評為 5* 級；其餘的則評為 5 級。表現低於第 1 級的水平定為「不予評級」，並不會在證書上匯報。

乙類科目

乙類應用學習科目的評核由提供相關課程的機構（課程提供者）負責。考生的成績須經考評局審定，並於香港中學文憑考試證書上匯報。由 2018 年文憑試開始，除應用學習中文課程（非華語學生適用）外，應用學習科目增至三個等級匯報成績，即「達標」、「達標並表現優異(I)」和「達標並表現優異(II)」。至於應用學習中文課程，則沿用現行的兩個等級匯報成績，即「達標」和「達標並表現優異」。表現低於「達標」的水平會被定為「不達標」，並不會在證書上匯報。

丙類科目

丙類科目包括六個其他語言科目，採用劍橋大學國際考評部（劍橋國際）的高級補充程度試卷，閱卷及成績評級亦由劍橋國際處理。有關科目的成績以五個等級（a-e）在香港中學文憑考試證書上匯報。在五個等級中，e 級最低，a 級最高，而 e 級以下的成績會定為「不予評級」，並不會在證書上匯報。

個別科目的特殊安排

中國語文和英國語文

除了科目的整體等級外，兩個語文科目均會匯報每個分部（例如閱讀）的等級。個別分部的成績若低於第 1 級，會定為「不予評級」。如考生的科目整體成績為第 1 級或以上，而有個別分部成績為「不予評級」，則所有分部和科目整體成績均會在香港中學文憑考試證書上匯報；若科目整體成績在第 1 級以下，則無論分部或整體科目成績，均不會在證書上匯報。

英國語文科的卷一閱讀和卷三聆聽及綜合能力考核均設有必答部分和選答部分。除必答部分外，考生可選擇作答較易或較難的部分。如考生選答較難部分，在相關分部最高可獲第 5** 級，選答較易部分的考生則最高只能獲得第 4 級。

數學

考生可以選擇只應考必修部分，或應考必修部分及延伸部分兩個單元的其中之一。必修部分及延伸部分的成績會分別匯報。

組合科學

本科的整體等級及兩個選修部分（生物、化學或物理）的等級均會在香港中學文憑考試證書上匯報。如考生的科目成績在第 1 級或以上而其中一分部成績為「不予評級」，則所有分部和科目整體成績均會在證書上匯報；若科目成績在第 1 級以下，則無論分部或整體科目的成績，均不會在證書上匯報。

3 等級描述

在水平參照成績匯報制度下，各科均附有一套等級描述，以說明該學科達至某等級的考生之典型能力水平。在擬訂等級描述時，以正面描述考生能展示的水平為原則，避免描述考生不能掌握的能力。

這些等級描述只是「一般性」的敘述，未必能準確描述個別考生的實際表現。這是由於某個考生可能在同一科目的不同卷別有跨越多於一個等級的表現。在每科的評核過程中，這些不同的表現會被整合，以最配合考生表現的等級來匯報其成績。

除了各科的等級描述外，考評局也制定了一套涵蓋不同科目的共通等級描述（表一），簡略說明考生於不同等級表現的水平。共通等級描述是根據各學科共同的主要學習成果而擬訂，包括對課程內容的認識和理解，應用概念和技巧的能力，詮釋、分析、綜合和評價等高階能力，以及傳意能力。

在首屆文憑試推出之前，24 個甲類科目已擬訂第 1 至 5 級的等級描述，說明各級考生的典型表現，每科的等級描述是由一個工作小組所擬訂，該小組的成員來自不同界別，當中包括課程發展主任、校長、中學教師，以及大學的學科專家。因應高中課程和評核的檢討，個別科目的評核有所改變，同時在審視考生真實表現的樣本後，我們於 2014 年整體檢視了 24 個科目的等級描述，個別科目也因應情況更新其等級描述。

文憑試各科最新版本的等級描述及考生表現示例，請參閱考評局網頁：

www.hkeaa.edu.hk/tc/hkdse/assessment/subject_information/

等級描述的用途

等級描述是科目專家在評級、設定水平和維持水平等工作的重要參考資料。有關詳情，可參考第 4 至 6 章。

同時，等級描述亦有助促進學與教。學生可參考有關描述，以制定個人的學習目標和評估學習進度；而教師亦可藉此明確瞭解如何協助學生邁向更高水平。

此外，對大專院校或僱主等持份者而言，等級描述讓他們更清楚瞭解不同水平考生的實際能力，從而作出更合適的甄選決定。

表一 共通等級描述

第 5 級

達到本等級考生的典型表現如下：

- 對課程內容有廣泛的認識和透徹的理解，能把概念和技巧有效地應用到多元和複雜的不熟悉情境，並顯示深入的見解
- 能分析、綜合和評價廣泛的資料
- 能精簡及邏輯地傳達意念和見解

第 4 級

達到本等級考生的典型表現如下：

- 對課程內容有良好的認識和理解，能把概念和技巧有效地應用到不熟悉的情境，並顯示深入的見解
- 能分析、綜合和詮釋各種資料
- 能邏輯地傳達意念和見解

第 3 級

達到本等級考生的典型表現如下：

- 對課程內容有足夠的認識和理解，能把概念和技巧適當地應用到不同的熟悉情境
- 能分析和詮釋各種資料
- 能恰當地傳達意念和見解

第 2 級

達到本等級考生的典型表現如下：

- 對課程內容有基本的認識和理解，能把概念和技巧應用到熟悉的情境
- 能辨識和詮釋直接的資料
- 能平衡地傳達簡單意念

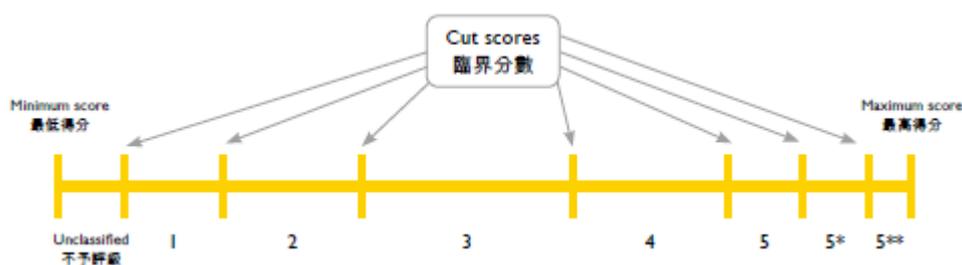
第 1 級

達到本等級考生的典型表現如下：

- 對課程內容有初步的認識和理解，在協助下，能把概念和技巧應用到簡單熟悉的情境
- 在指導下，能辨識和詮釋簡單的資料
- 能粗略地傳達簡單意念

4 評級程序

評級主要的目的是釐定考生須至少獲得多少分才達到某個等級。每個等級的最低分數稱為臨界分數（圖一）。



圖一 臨界分數

文憑試的評級程序涉及的一系列工作，在正式閱卷前已展開。每一科目均成立一個專家小組執行評級工作，小組成員包括科目經理、試卷主席、經甄選的助理試卷主席或相關的閱卷員。具體程序如下：

選取樣本答卷

考試完成後，小組成員將選取最能符合等級描述的答卷作為樣本答卷。

標準化

選取樣本答卷後，小組將討論樣本答卷的評分和具體得分點。他們也會比較當年與往年的水平，並留意兩者的差異。經評閱的樣本答卷會用作閱卷時參考的標準答卷。

閱卷後程序

完成閱卷和校本評核分數調整（如適用）後，小組會檢視公開考試和校本評核樣本的示例，並綜合觀察所得，討論各等級水平的要求。有關討論讓專家小組依據樣本的表現作出判斷，為各試卷或分部作初步的評級建議（包括初步臨界分數範圍）。有關設定水平和維持水平的詳情，可參考第5章和第6章。

專家小組評級會議

考評局評核發展部與評核科技及研究部管理小組的成員，會與專家小組商議，就各個科目及其分部的臨界分數達成共識。當專家建議的臨界分數，與統計分析結果建議的臨界分數出現明顯差異時，將會適當考慮專家的意見，而統計結果則作為參考，以驗證臨界分數的合理性。如未能達成共識，專家小組將全面討論不同的意見，並記錄其理據。專家小組可在會議中探討修訂各卷的臨界分數對整體等級分布的影響。最後，專家小組會建議有關科目的臨界分數。

內部會議

以考評局秘書長為首的高層管理小組，會在內部會議中審視專家小組的評級建議。如統計數據與專家小組建議出現分歧，並無法於專家小組評級會議取得共識，內部會議將適當地平衡統計數據與專家小組的建議，對有關科目的臨界分數作出凌駕性建議，供公開考試委員會審定。

公開考試委員會會議

公開考試委員會將審定內部會議的建議，並通過所有科目的最終臨界分數。

5 設定水平

所有甲類科目第 1 至 5 級的水平已於 2012 年首屆文憑試設定，並於 2013 年及往後年度的文憑試維持穩定，以確保各年度考試成績的可比性。

在眾多設定水平的方法中，考評局採用一套結合專家判斷和統計技術的方法。當中，專家會參考統計模型的建議及其他參考資料，釐定最合適的臨界分數。原則上，2012 年文憑試各科的水平按照等級描述和考生實際表現而設定。不同科目會利用不同的統計方法計算建議臨界分數，以供專家參考。有關統計方法如下：

核心科目

文憑試四個核心科目，即中國語文、英國語文、數學及通識教育，其水平設定參考 2012 年考生實際表現與以下各項的關係：

- 等級描述；
- 2011 年香港高級程度會考（高考）考生水平；
- 2010 年香港中學會考（會考）考生水平。

為了解文憑試與以往的高考和會考的水平，考評局分別邀請部分參與 2010 年會考和 2011 年高考的學校，進行保密研究測驗。同樣的研究測驗並在參與 2012 年文憑試的部分學校中進行。從研究測驗取得的統計數據，可作為連結首屆文憑試和以往高考與會考水平的參考資料。

專家小組在釐定臨界分數時，也會參考以下資料：

- 按總分選取的已評閱答卷；
- 該年試卷間的相關係數、平均分及標準差；
- 試卷總分和科目總分的累積分布；
- 閱卷員就個別試卷難易程度的回饋意見；
- 水平參照成績匯報資料套內不同能力水平 的樣本；
- 2010 年會考和 2011 年高考的試卷和存檔 的參考答卷。

選修科目

考評局採用「組別能力指數」作為選修科目評級的參考。「組別能力指數」是以統計方法計算的一組百分比，以得出建議臨界分數。在計算「組別能力指數」時，會依據應考有關選修科目的考生在核心科目取得某等級的人數，來計算出該選修科於相應等級的「組別能力指數」。有關「組別能力指數」的計算方法詳列於附錄 1。

在 2012 年首屆文憑試，「組別能力指數」建議的臨界分數只用作參考，各科的專家小組主要按考生的實際表現，獨立為各級水平釐定合適的臨界分數。

應用學習科目

課程提供者先設定「達標」的水平，再由專家小組按個別科目的表現描述作調整。從 2018 年文憑試開始，「達標並表現優異(I)」和「達標並表現優異(II)」的水平分別等同甲類科目的第 3 級和第 4 級或以上，因此獲得「達標並表現優異(I)」和「達標並表現優異(II)」等級的考生比例，會參考「組別能力指數」第 3 級和第 4 級或以上的百分比來釐定。

至於應用學習中文課程的「達標」及「達標並表現優異」均採用專家判斷方法設定水平。

個別科目的特殊安排

文憑試部分科目設有選修部分。為劃一不同選修部分考生的水平，考評局採用百分位等值法為不同選修部分作分數等值。該方法的原理和應用載於附錄 2。

5**級與 5*級的釐定

5**級和 5*級的臨界分數是按第 5 級內考生的分數分布而訂定。其中，約 10%成績最優異的第 5 級考生，可獲 5**級，約 30%成績次佳的第 5 級考生，可獲 5*級。

至於中國語文和英國語文科，專家小組會先釐定整個科目 5**級和 5*級的臨界分數，然後推算和釐定個別分部的臨界分數，使各分部的 5**級與 5*級的分布較為一致；然而，各分部的臨界分數相加後，仍須相等於整個科目相應等級的臨界分數。

就組合科學而言，生物、化學及物理等分部的臨界分數，會採用百分位等值法，根據相應的三個科目的臨界分數來推算，並參考各分部科目的「組別能力指數」，以及考生的實際表現而釐定。組合科學的分部科目的 5**級和 5*級臨界分數，會按其第 5 級的分數分布，根據上述百分比釐定。組合科學整科的臨界分數，則是把有關的兩個分部科目的臨界分數相加而成。故此，在組合科學科考獲第 5 級的考生當中，能取得 5**級和 5*級的人數比例並不一定如其他科目般定為 10%和 30%。

6 維持水平

與水平設定的程序相若，不同科目的專家小組會參考各種資料以釐定臨界分數。這些資料包括：

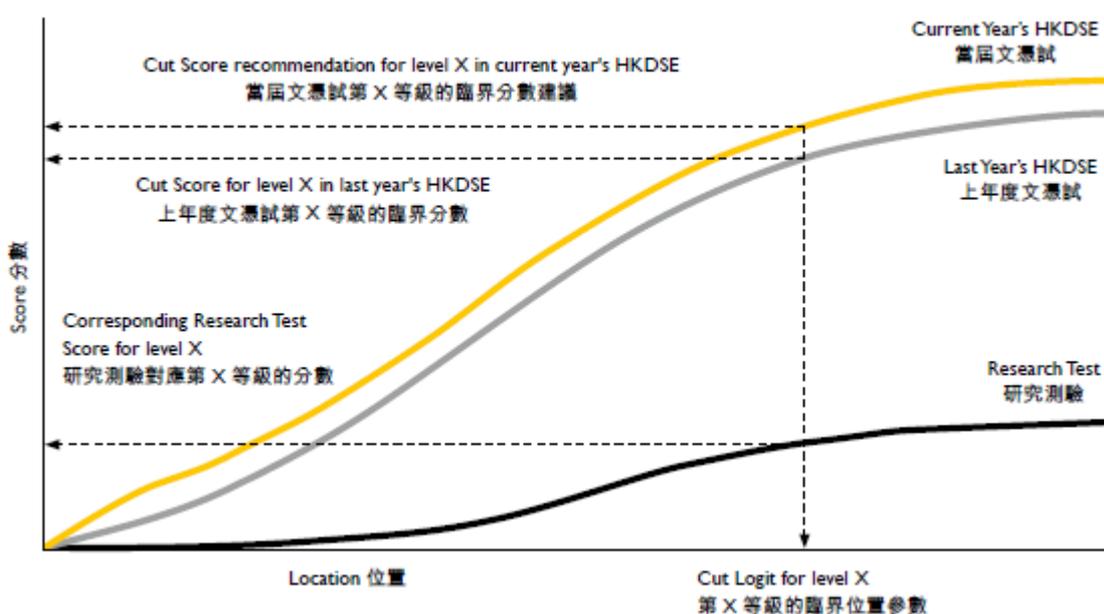
- 按總分選取的已評閱答卷；
- 該年試卷間的相關係數、平均分及標準差 與往年的比較；
- 往年各分題的統計數據；
- 試卷總分和科目總分的累積分布；
- 閱卷員就個別試卷難易程度的回饋意見；
- 文憑試存檔的參考答卷。

專家小組亦會參考統計模型建議的臨界分數。核心科目和其他科目會採用不同統計模型來計算建議臨界分數。

核心科目

考評局每年均會小心選取一些學校，為四個核心科目進行保密研究測驗。根據研究測驗與該年公開考試的數據，運用「潛在特性模型」校定不同考試的題目的難度，並以此計算建議臨界分數。有關「潛在特性模型」的詳情，可參考附錄 3。

如圖二所示，運用研究測驗結果，通過等值方法，將當屆文憑試和上一年度的考試等值。而根據上一年度考試設定的臨界分數，找到當屆考試的等值對應分數，就是統計建議的臨界分數。這樣，專家小組可參考年與年之間核心科目考試成績的等值結果，再根據第 4 章的程序，釐定最終的建議臨界分數。



圖二 利用研究測驗來維持核心科目水平及產生參考臨界分數

選修科目

選修科目的建議臨界分數是依據「組別能力指數」計算的。與核心科目的評級方法相若，選修科目的專家小組可參考以「組別能力指數」計算所得的建議臨界分數，並藉此作最終的建議，以維持各等級的水平。

應用學習科目

為了維持「達標」水平，考評局會參照個別科目的表現描述和往年考生的真實表現。考生達到「達標並表現優異(I)」和「達標並表現優異(II)」等級的比例，則參照該組別考生於核心科目中達第3級和第4級或以上的「組別能力指數」而釐定。

其他語言科目

由於這些科目的考試採用劍橋大學國際考評部高級補充程度的試卷，故有關科目的水平亦由該機構維持。

以上維持水平的方法可以向公眾、香港及海外的大專院校、僱主及其他使用者，確保各等級的水平標準是固定的，日後亦不會出現「等級膨脹」的情況。

7 匯報成績

為提供更詳細的資料，香港中學文憑考試證書列出科目及分部（如有）的成績等級。校本評核是甲類科目不可分割的部分，因此校本評核的成績不會獨立匯報。經過調整的校本評核分數（如有）將與公開考試成績合併，並在分部或科目等級中一併匯報。香港中學文憑考試證書如圖三所示。

香港考試及評核局
HONG KONG EXAMINATIONS AND ASSESSMENT AUTHORITY
香港中學文憑考試
HONG KONG DIPLOMA OF SECONDARY EDUCATION EXAMINATION

茲證明
This is to certify that

陳大文
CHAN TAI MAN

參加 2018 年香港中學文憑考試，成績如下：
sat the 2018 Hong Kong Diploma of Secondary Education Examination and achieved the following results:

科目 Subject	科目等級 Subject Level / Grade	分部等級 Component Level
甲類學科 Category A Subjects		
中國語文	CHINESE LANGUAGE	5(Five)
閱讀	Reading	5(Five)
寫作	Writing	4(Four)
聆聽與綜合能力	Listening and Integrated Skills	5(Five)
說話	Speaking	3(Three)
英語語文	ENGLISH LANGUAGE	3(Three)
閱讀	Reading	3(Three)
寫作	Writing	2(Two)
聆聽與綜合能力	Listening and Integrated Skills	3(Three)
說話	Speaking	3(Three)
數學	MATHEMATICS	4(Four) & 3(Three)
必修部分及延伸部分（包括必修部分）	Compulsory Part & Extended Part (Mandatory and Elective)	
通識教育	LIBERAL STUDIES	5(Five)
企業、會計與財務概論（會計）	BUSINESS, ACCOUNTING AND FINANCIAL STUDIES (ACCOUNTING)	5(Five)
乙類學科 Category B Subjects		
形象設計	IMAGE DESIGN	達標 Attained
丙類學科 Category C Subjects		
日語	JAPANESE LANGUAGE	3(B)
閱讀能力、寫作	With Reading & Speaking	
科目總數 Total Number of Subjects	SEVEN	
身分證照文件號碼 Identification Document No.	Z900006(9)	考生編號 Candidate No.: 18000006

註釋及附頁
Explanatory Notes are printed overleaf

證書編號
Certificate No.: 18000001


 蘇國生 SO Kwok Sang
 秘書長 Secretary General

圖三 香港中學文憑考試證書樣本

附錄 1：組別能力指數

「組別能力指數」基本上可視作一組百分比，用作選修科目和應用學習科目評級的參考。

計算某科目 X 某個等級或以上（例如第 3 級或以上）的組別能力指數 P 的公式如下：

$$P = \frac{1}{b_C + b_E + b_M + b_L} \left(b_C \frac{n_C}{N_C} + b_E \frac{n_E}{N_E} + b_M \frac{n_M}{N_M} + b_L \frac{n_L}{N_L} \right) \times 100\%$$

其中 N_C 、 N_E 、 N_M 及 N_L 分別代表科目 X 的考生應考四個核心科目：中國語文、英國語文、數學（必修部分）及通識教育科的總人數；

n_C 、 n_E 、 n_M 及 n_L 分別代表應考某個科目 X 的考生在中國語文、英國語文、數學（必修部分）及通識教育科中獲得某個等級或以上（例如第 3 級或以上）的人數；

b_C 、 b_E 、 b_M 及 b_L 是把科目 X 的標準分與四個核心科目的標準分作回歸分析，得出的回歸係數。

計算出 P 的數值後，某個等級或以上（例如第 3 級或以上）的建議臨界分數為科目 X 分數的 $1-P$ 百分位點。

附錄 2：百分位等值法

部分文憑試的試卷包括一個必答部分，和兩個或多個選答部分。因此，考評局需要進行分數等值，使選答不同部分考生的表現，能以同一尺度作比較。

分數等值的概念是把某選答部分的分數，以必答部分的分數作中介，轉換為另一選答部分的分數。又或是把不同選答部分的分數都轉換為必答部分的分數。分數等值涉及三個步驟：

- 按考生表現把每個部分的分數排序；
- 把每一部分的考生分為多個等份（例如分 20 等份，每等份佔全部考生 5%）；
- 按各等份的分數分布，把某一部分的分數 轉換為另一部分的分數；如有需要調整相同百分位點的分數。

以英國語文科為例，公開考試部分包括四份試卷：分別是閱讀（卷一）、寫作（卷二）、聆聽及綜合能力考核（卷三）與說話（卷四）。寫作和說話卷中所有部分均屬必答。而閱讀和聆聽及綜合能力考核兩卷，則只有 A 部分為必答題，考生須作答 A 部分和選答 B1 部分（較易部分）或 B2 部分（較難部分）的題目。

為了以劃一的尺度評估選答不同部分考生的表現，考評局將參考必答部分的分數分布，把較易部分的分數轉換到一個與較難部分可比的尺度。轉換後的結果稱為較難部分的等值分數。把每一考生必答部分的分數和較難部分的分數（或較難部分的等值分數）相加，便得出試卷總分，繼而根據試卷總分進行評級。

附錄 3：潛在特性模型

「潛在特性模型」(latent trait model) 是建基於現代題目反應理論的統計工具，能通過分析考生對題目的反應（考試表現）校定考試題目及估計學生的能力水平。

校定文憑試題目的「潛在特性模型」是根據 Rasch 模型發展出來的。簡單的二值（即只有「對」或「錯」兩個答案）反應 Rasch 模型如下：

$$p_{ni} = \Pr\{X_{ni} = 1\} = \frac{\exp(\beta_n - \delta_i)}{1 + \exp(\beta_n - \delta_i)} \quad (1)$$

但是，大部分考試題目都不是簡單的只有「對」或「錯」兩個答案的問題，每題答對的得分亦可能多於一分。這時，需要以下的多級評分模型：

$$\Pr\{X_{ni} = x_{ni}\} = \frac{\exp\{x_{ni}(\beta_n - \delta_i) - \sum_{k=0}^{x_{ni}} \tau_{ik}\}}{\sum_{l=0}^{m_i} \exp\{l(\beta_n - \delta_i) - \sum_{k=0}^l \tau_{ik}\}} \quad (2)$$

其中 β_n 是考生 n 的能力， δ_i 是題目 i 的難度， $\{\tau_{ik}\}$ 是已經中心化的閾值 (centralised thresholds)，

而 $\sum_{k=0}^0 \tau_{ik} = 0$ 。 β_n 與 δ_i 則統稱為位置參數(logits)。

這是考評局用作校定某科目公開考試題目與研究測驗題目的模型。為把不同年份的考試分數等值，考生在某科目任何題目子集的總分將按以下公式計算：

$$SubS = \sum_{i \in I} K_i E(X_i) \quad (3)$$

其中 K_i 是題目 i 的設定權重， I 是該科目的題目子集，而 $E(X_i)$ 是某一能力水平為 β 的考生 n 在某一滿分為 m_i 分的題目 i 的預期分數：

$$E(X_i) = \sum_{x'=1}^{m_i} x' P(X_i = x' | \beta) \quad (4)$$

通過同時校定各份試卷和保密研究測驗的所有題目，便可以用相同尺度估計每個考生的能力。

當校定試題的難度後，便可為某一考生能力值計算相對應的整個科目（亦可能包括不同分部）的分數期望值。在計算過程中，會顧及取得某個分數的概率，以及科目考試可能包括多個獨立分部，每一分部有不同的比重。該比重是根據試卷結構及考生表現等資料計算出來的。

本模型產生四個核心科目各等級的建議臨界分數，供專家小組評級時參考。

從 2013 年開始，為提供一個能將連續多個年度考試等值的基本參考資料，考評局採用以上的「潛在特性模型」校定（某兩年）公開考試每份試卷中所有題目與保密研究測驗中所有題目的參數，以作為維持每年等級水平的參考。

如有查詢，請聯絡：

香港考試及評核局

地址：香港灣仔軒尼詩道 130 修頓中心 13 樓

電話：(852) 3628 8860

傳真：(852) 3628 8928

電郵：dse@hkeaa.edu.hk