

Grading Procedures and Standards-referenced Reporting in the HKDSE
March 2018

Table of Contents

1. Introduction

2. The reporting system

3. Level descriptors

4. The grading procedures

5. Setting the standards

6. Maintaining the standards

7. Reporting the results

Appendix 1: Group Ability Index

Appendix 2: Equipercentile method

Appendix 3: Latent trait model

1. Introduction

In the Hong Kong Diploma of Secondary Education Examination (HKDSE), Standards-referenced Reporting (SRR) is adopted to report candidates' assessment results. For a given subject, candidates' levels of performance are reported with reference to a set of explicit and fixed standards of performance.

SRR is not a new approach in public examinations in Hong Kong. It has been in use for the subjects of Chinese Language and English Language in the Hong Kong Certificate of Education Examination (HKCEE) since 2007. This step was undertaken to ensure that results are reported with reference to standards that are internationally recognised, transparent, explicit, and that remain constant over time.

To cater for the SRR approach, the Hong Kong Examinations and Assessment Authority (HKEAA) has established a grading system which aims to maintain standards between successive years and inter-subject comparability within the same year.

The purpose of this booklet is to explain the rationale, approaches and key features of the grading procedures and the SRR approach.

2 The reporting system

There are three categories of HKDSE subjects: Categories A, B and C.

Category A subjects

The HKEAA is responsible for the grading of all Category A Senior Secondary subjects. Candidates' performance is reported on the HKDSE certificate at five levels (1–5), with 5 being the highest. The highest-achieving level 5 candidates are awarded level 5**; the next highest-achieving level 5 candidates are awarded level 5* and the remaining level 5 candidates receive simply level 5. Performance below the level 1 standard for a subject is designated as 'Unclassified'. This will not be reported on the certificate.

Category B subjects

The assessment of Category B Applied Learning (ApL) subjects is undertaken by the organisations providing relevant courses (course providers). After moderation by the HKEAA, the final results are recorded on the HKDSE certificate. Starting from the 2018 HKDSE, an additional level of performance is introduced to ApL subjects, except for ApL Chinese (for non-Chinese speaking students). Three levels of performance are reported, namely 'Attained', 'Attained with Distinction (I)' and 'Attained with Distinction (II)'. For ApL Chinese, the existing levels of performance, i.e. 'Attained' and 'Attained with Distinction', continue to be used. Performance below 'Attained' is designated as 'Unattained' and not reported on the certificate.

Category C subjects

Category C subjects comprise six Other Language subjects. Question papers of the Advanced Subsidiary (AS) level from Cambridge Assessment International Education (Cambridge International) are used. Marking and grading are conducted by Cambridge International. Results are reported in five grades (a-e) on the HKDSE certificate, with grade e being the lowest and grade a being the highest. Achievements below grade e are designated as 'Ungraded' and not reported on the certificate.

Special arrangements for individual subjects

Chinese Language and English Language

In addition to the overall subject level, candidates' performance in individual components (e.g. Reading) is also reported. Performance below the level 1 standard for individual components is designated as 'Unclassified'. Candidates whose performance is designated as 'Unclassified' in some components but who achieve level 1 or above at subject level will have all component and subject levels reported on the HKDSE certificate. However, if the subject level is below level 1, neither the subject level nor individual component levels will be reported on the certificate.

In English Language, Paper 1 Reading and Paper 3 Listening and Integrated Skills both comprise a compulsory section and a section where candidates can choose between an easier or a more difficult option. Candidates attempting the more difficult option can attain up to level 5** in the relevant

component, while those who attempt the easier option can only attain up to level 4.

Mathematics

A candidate may take the Compulsory Part only or take both the Compulsory Part and one of the two modules in the Extended Part. The results of the Compulsory Part and the Extended Part are reported separately.

Combined Science

The overall subject level is reported together with the levels for the two chosen components (Biology, Chemistry, or Physics). Candidates whose performance is designated as 'Unclassified' in one component but achieve level 1 or above at the subject level will have all component and subject levels reported on the HKDSE certificate. However, if the subject level is below level 1, neither the subject level nor individual component levels will be reported on the certificate.

3 Level descriptors

Under the Standards-referenced Reporting system, a set of descriptors is written for each subject to describe what a candidate performing at a given level is typically able to do. The principle used in preparing these descriptors is to describe the achievements that candidates are able to demonstrate in positive terms and to avoid describing what candidates cannot achieve.

By their nature, these descriptors represent 'on-average' statements, and may not precisely describe the actual performance of individual candidates. A given candidate may demonstrate, across the examination papers for a single subject, some elements of performance that are characteristic of more than one level. The assessment process that is applied within each subject aggregates and combines such variation to derive a final result that provides a 'best fit' between a given candidate's performance and a level descriptor.

In addition to the subject level descriptors, a set of generic descriptors has been developed to provide an overall brief description of the performance standards of candidates at different levels across subjects (Table 1). The generic descriptors encompass the major learning outcomes that are common across subjects, including knowledge and understanding of the curriculum, the ability to apply concepts and skills, higher order abilities such as interpretation, analysis, synthesis and evaluation, and the ability to communicate.

Before the launch of the HKDSE, level descriptors depicting the typical performance standard of candidates at levels 1 to 5 of individual subjects had been developed for the 24 Category A subjects by working groups that comprised members from different sectors, including officers from the Curriculum Development Institute (CDI), school principals, secondary school teachers, and university subject experts. With the changes in the assessment of a number of subjects as a result of the reviews of Senior Secondary curriculum and assessment, and after reviewing samples of actual performance in the live examinations, an overall review of the level descriptors of all 24 subjects were conducted in 2014 and the descriptors for individual subjects were updated as appropriate.

The latest version of the level descriptors and samples of candidates' performance of HKDSE subjects are available at the HKEAA website: www.hkeaa.edu.hk/en/hkdse/assessment/subject_information/

Use of level descriptors

Level descriptors are important sources of reference for subject experts to make judgments in grading, standards setting and maintenance. For details, please refer to sections 4 to 6.

The level descriptors also facilitate learning and teaching. Students can make use of the descriptors to set their goals and assess their learning progress. Teachers also know explicitly what they have to do to assist students to achieve a higher level of performance.

Moreover, the level descriptors are useful for other users, such as tertiary institutions or employers, in enabling them to better understand what candidates at different levels of performance can do in order for them to make more appropriate selection decisions.

Table 1 Generic descriptors

Level 5

Candidates at this level typically demonstrate:

- comprehensive knowledge and understanding of the curriculum and the ability to apply the concepts and skills effectively in diverse and complex unfamiliar situations with insight
- ability to analyse, synthesise and evaluate information from a wide variety of sources
- ability to communicate ideas and express views concisely and logically

Level 4

Candidates at this level typically demonstrate:

- good knowledge and understanding of the curriculum and the ability to apply the concepts and skills effectively in unfamiliar situations with insight
- ability to analyse, synthesise and interpret information from a variety of sources
- ability to communicate ideas and express views logically

Level 3

Candidates at this level typically demonstrate:

- adequate knowledge and understanding of the curriculum and the ability to apply the concepts and skills appropriately in different familiar situations
- ability to analyse and interpret information from a variety of sources
- ability to communicate ideas and express views appropriately

Level 2

Candidates at this level typically demonstrate:

- basic knowledge and understanding of the curriculum and the ability to apply the concepts and skills in familiar situations
- ability to identify and interpret information from straightforward sources
- ability to communicate simple ideas in a balanced way

Level 1

Candidates at this level typically demonstrate:

- elementary knowledge and understanding of the curriculum and the ability to apply the concepts and skills in simple familiar situations with support
- ability to identify and interpret information from simple sources with guidance
- ability to communicate simple ideas briefly

4 The grading procedures

The main purpose of grading is to determine the minimum score needed for a candidate to attain a given level. This minimum score is known as the cut score (Figure 1).

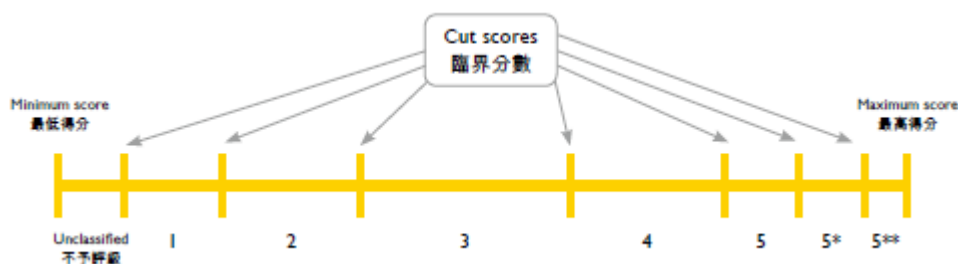


Figure 1 Cut scores

The HKDSE grading procedures include a series of tasks that begins before the actual marking of scripts. For any given subject, a panel of expert judges consisting of the subject manager(s), the chief examiners and selected assistant examiners or markers from the individual components is responsible for the implementation of the following grading procedures.

Sample script selection

After the examination, some samples that panel members feel illustrate performance particularly well in relation to the level descriptors will be selected.

Standardisation

After script selection, the panel will discuss the marks to be awarded to discrete points in the sample scripts. They will also make observations on any perceived changes in standard from the previous year's work. These marked scripts will be used as standardisation scripts for marking.

Post-marking exercise

After completion of marking and moderation of School-based Assessment (SBA) scores (if appropriate), the panel will discuss the standard requirement of each level based on exemplars and/or observations from the live examination and SBA sample review. The objective of the discussion is to make provisional grading recommendations (including preliminary cut score ranges) on each examination paper or component through expert judgment based on samples of performance. Details regarding standards setting and maintenance are explained in sections 5 and 6.

Panel of judges grading meeting

Management team members from the Assessment Development Division and the Assessment Technology and Research Division of the HKEAA will meet with the panel judges and agree on cut scores for each paper/component and for the subject. If there is any discrepancy between their recommendations and the statistical indicators, the panel's expert recommendations will be given appropriate weighting in this exercise. The statistical indicators should, therefore, be used to check and verify the reasonableness of the cut scores proposed by the panel of judges for each subject. Any

significant discrepancies between the panel's recommendations and the statistical indicators should be fully discussed and if they cannot be resolved, the rationale behind the panel's recommendations should be presented. During this meeting, the panel of judges can investigate the impact of amending the cut scores for each examination paper on subject grade distributions. By the end of the meeting for each subject, the panel of judges will decide on their recommendations for the cut scores for that subject.

Internal meeting

A senior management team led by the Secretary General of the Authority will meet to review the recommendations made by individual subject panels. In cases where significant discrepancies remain between the recommendations of the judge panel and the statistical indicators after the panel of judges grading meetings, this internal meeting will make overriding recommendations to the Public Examinations Board on the appropriate alignment between statistical indicators and the subject experts' recommendations for each subject.

Public Examinations Board meeting

The recommendations of the internal meeting will be reviewed at the Public Examinations Board meeting. The final cut scores of all HKDSE subjects must be endorsed at this meeting.

5 Setting the standards

The standards of levels 1 to 5 of all Category A subjects were set in 2012 HKDSE. The standards are maintained in subsequent years starting from 2013 so that results are comparable across years.

There are various methods that can be used to set standards. The HKEAA adopts a strategy that combines the benefits of both expert judgment and statistical methods. Experts make use of statistical recommendations and other reference data to make the most robust decisions possible in cut score determination. In principle, the standards of all HKDSE subjects in the 2012 examination were set with reference to the level descriptors and actual performance of candidates in the examination. To support experts' decision making, different statistical techniques have been applied to different subjects to produce recommended cut scores. The methodologies are as follows:

Core subjects

In setting standards for the four HKDSE core subjects - Chinese Language, English Language, Mathematics and Liberal Studies, reference were made to the actual performance of candidates in the 2012 HKDSE in relation to:

- the level descriptors;
- the standards of candidates in the 2011 Hong Kong Advanced Level Examination (HKALE);
- the standards of candidates in the 2010 Hong Kong Certificate of Education Examination (HKCEE).

In order to provide some background information about standards in the HKDSE and previous HKCEE/HKALE, secure research tests for the core subjects were administered to samples of HKCEE school candidates in 2010 and HKALE school candidates in 2011. The same research tests were also administered to a selected sample of candidates sitting the 2012 HKDSE. The statistical data from the research tests provided some linkage of the standards in the HKDSE and the previous HKCEE and HKALE.

The panel of judges will refer to the following information in making their decisions on cut scores:

- marked live scripts, selected according to total marks;
- inter-paper correlations, the mean and standard deviation of the current year's papers;
- paper mark and subject mark cumulative distributions;
- feedback from markers on the level of difficulty of each particular examination paper;
- performance samples from the HKDSE SRR Information Packages;
- HKALE and HKCEE papers and library scripts from 2011 and 2010 examinations respectively.

Elective subjects

A Group Ability Index (GAI) is used as a reference for grading elective subjects. The GAI is a set of percentages generated statistically for obtaining a suggested set of cut scores. For a group of candidates taking a specific elective subject, a GAI can be calculated for each level based on the number of candidates in this group achieving that level in the four core subjects. The method of GAI

calculation is detailed in Appendix 1.

In the first administration of the HKDSE in 2012, the GAI-based cut scores were used as reference only. The panel of judges for each subject played a very important and independent role in setting a cut score for each level based on the actual performance of candidates.

Applied Learning subjects

Standards of 'Attained' are initially determined by course providers and moderated by panels of judges with reference to the performance descriptors of individual subjects. Starting from 2018 HKDSE, candidates awarded 'Attained with Distinction (I)' and 'Attained with Distinction (II)' are deemed to have performed at a level comparable to level 3 and level 4 or above respectively for Category A subjects. Therefore, the award percentages of 'Attained with Distinction (I)' and 'Attained with Distinction (II)' are determined based on the GAI for level 3 and level 4 or above.

As for ApL Chinese, both 'Attained' and 'Attained with Distinction' are determined by expert judgment.

Special arrangements for individual subjects

Several HKDSE subjects consist of elective modules/parts. In order to set common standards for candidates choosing different modules/parts of the same subject, it is necessary to equate marks in different modules/parts. A method of equipercetile mark conversion will be used for this purpose. The principle and application of this method are explained in Appendix 2.

Determination of level 5 and level 5***

The cut scores for level 5** and level 5* are set with reference to the percentage in mark distribution so that level 5** is awarded to the highest achieving 10% (approximately) of level 5 candidates and level 5* is awarded to the next highest-achieving 30% (approximately) of level 5 candidates.

For Chinese Language and English Language, the overall subject cut score for level 5** and level 5* are set first. The corresponding cut scores for each component are then determined by exploration and negotiation across the components in such a way that they represent a similar profile of division within level 5 for each component. However, the cut scores established for each component still aggregate to achieve the correct corresponding subject cut scores for level 5** and level 5*.

For Combined Science, the cut scores of the individual components of Biology, Chemistry and Physics are determined with reference to the cut scores of the corresponding subjects using the equipercetile method, the GAI of each half subject and the actual performance of candidates in the live examination. The cut score for level 5** and level 5* for each half subject are determined by applying the above-mentioned percentages to level 5 candidates. The half subject cut scores are then aggregated to obtain the subject cut scores for each subject pairing of Combined Science. In other words, the respective percentage of level 5 candidates attaining subject levels 5** and 5* in Combined Science may not necessarily be 10% and 30% as in other subjects.

6 Maintaining the standards

Similar to the standards setting procedures, panel judges of different subjects refer to various kinds of information for the determination of cut scores, including:

- marked live scripts, selected according to total marks;
- inter-paper correlations, the mean and standard deviation of the current year's papers in comparison to previous years;
- item statistics for previous years for reference;
- paper mark and subject mark cumulative distributions;
- feedback from markers on the level of difficulty of each particular examination paper;
- HKDSE library scripts.

The panel judges also refer to results generated by statistical models. The statistically-generated suggested cut scores for the core subjects and those for the other subjects are computed using different methods.

Core subjects

For the four core subjects, secure research tests are conducted annually in a representative group of carefully selected schools. Using both research test data and live examination data as input, a latent trait model is used to calibrate all questions in different examinations and generate suggested cut scores. The details of a latent trait model are documented in Appendix 3.

As shown in Figure 2, the statistically-generated suggested cut scores for the current year's examinations are obtained by mapping the cut scores established for last year's examinations to the live examination data with the standard maintained by means of test equating techniques using the results from the research tests. In this way, the statistical equating of successive examinations in the core subjects will support the panel judges in making their final decisions on the recommended cut scores, using the grading procedures described in section 4.

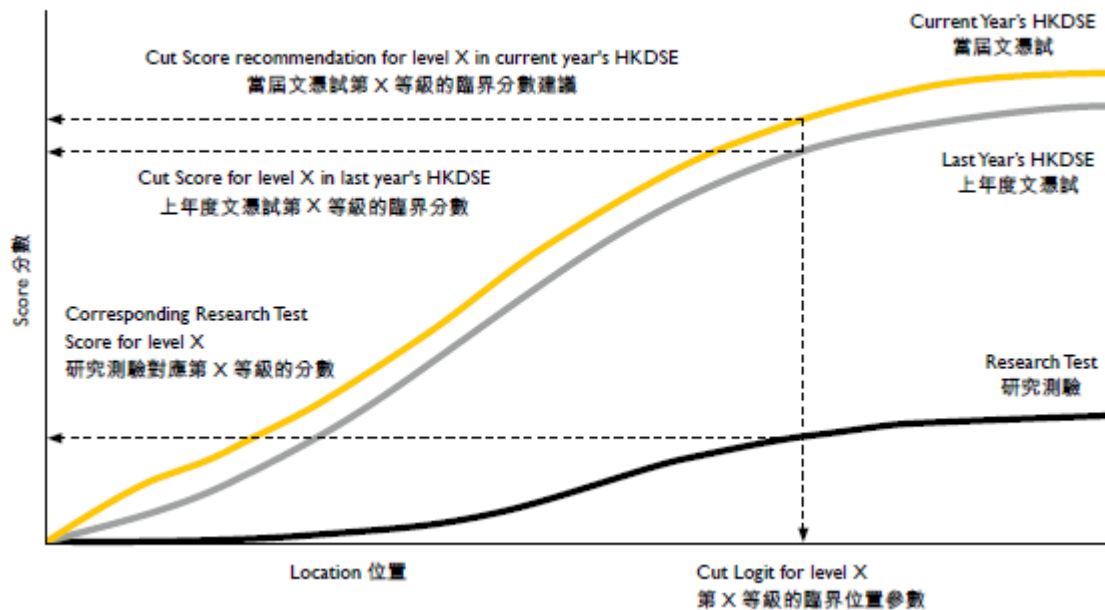


Figure 2 Using research tests in maintaining standards and producing reference cut scores for the core subjects

Elective subjects

Recommended cut scores for the elective subjects are suggested using the GAI. Similar to the grading of the core subjects, panel judges for the elective subjects may then consider these suggested cut scores and use them to support their final recommendations on standards-maintenance.

Applied Learning subjects

The standards of 'Attained' are maintained by making reference to the performance descriptors of individual subjects and the actual performance of candidates in the previous years. For 'Attained with Distinction(I)' and 'Attained with Distinction(II)', the GAI is used to determine the proportion of candidates to be awarded the corresponding levels with reference to the same group of candidates who have attained level 3 and level 4 or above in the core subjects.

Other Language subjects

The standards of these subjects are maintained by Cambridge Assessment International Education since their AS level papers are used for the examinations of these subjects.

The above standards-maintenance processes can assure the public, local and overseas tertiary institutions, employers, and other users that the standards are held constant and there is no 'grade inflation' over time.

7 Reporting the results

In order to provide more detailed information, the HKDSE certificate lists both subject level results as well as component level results (if any). SBA is an integral part of Category A subjects, so SBA results are not reported separately. Moderated SBA results (if any) are combined with public examination results to form an overall component or subject level as appropriate. An example of how results are set out on the HKDSE certificate is given in Figure 3.

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	分級等級 Component Level
甲類學科 Category A Subjects		
中國語文 閱讀 寫作 聆聽與綜合能力 說話	CHINESE LANGUAGE Reading Writing Listening and Integrated Skills Speaking	5(Five)
英語語文 閱讀 寫作 聆聽與綜合能力 說話	ENGLISH LANGUAGE Reading Writing Listening and Integrated Skills Speaking	3(Three)
數學 必修部分及延伸部分（數學分析與統計）	MATHEMATICS Compulsory Part & Extended Part (Calculus and Statistics)	4(Four) & 3(Three)
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	成績 Attained
丙類學科 Category C Subjects		
日語 閱讀力：寫字	JAPANESE LANGUAGE With Reading & Writing	B0
科目總數 Total Number of Subjects	七 SEVEN	
身分證件號碼 Identification Document No.	29000069	考生編號 Candidate No.: 18000006

註釋及說明
Explanatory Notes are printed overleaf

證書編號
Certificate No.: 18000001


 蔡國生 SO Kwok Sang
 秘書長 Secretary General

Figure 3 Sample of HKDSE certificate

Appendix 1: Group Ability Index (GAI)

Group Ability Index (GAI) can be regarded as a set of 'suggested' percentages that are used as a reference for the grading of elective subjects and Applied Learning subjects.

The formula for calculating the Group Ability Index P of Subject X for a certain level or above (e.g. level 3 or above) is as follows:

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

where N_C , N_E , N_M and N_L are the numbers of Subject X candidates who sat the four core subjects - Chinese Language, English Language, Mathematics (Compulsory Part) and Liberal Studies, respectively;

n_C , n_E , n_M and n_L are the numbers of Subject X candidates who obtained that level or above (i.e. level 3 or above) in the four core subjects of Chinese Language, English Language, Mathematics (Compulsory Part) and Liberal Studies, respectively;

b_C , b_E , b_M and b_L are coefficients obtained by regressing the standardised scores of Subject X on the standardised scores of the four core subjects.

After calculating the value of P , the suggested cut score for that level or above (i.e. level 3 or above) is the $1-P$ percentile of the Subject X scores.

Appendix 2 Equipercentile method

In some examination papers of HKDSE subjects, there are one compulsory part and two or more optional parts. Equating is needed so that the performance of candidates choosing different optional parts can be reflected on the same scale.

The idea of equating is to convert the marks of one optional part into another optional part by using the compulsory part as a mediator, or to convert marks of all optional parts into the marks of the compulsory part. This can be done in three steps:

- Ranking the candidates according to their performance in each part;
- Dividing the candidates of each part into equal-sized groups (for example 20 groups of candidates, each made up of 5% of the total candidature);
- Converting the marks of one part to the marks of another part with reference to the mark distribution in different groups; if necessary, making adjustments in cases with the same percentile.

For example, in the subject of English Language, there are four papers in the public examination, namely Reading (Paper 1), Writing (Paper 2), Listening and Integrated Skills (Paper 3) and Speaking (Paper 4). All parts in Writing and Speaking are compulsory. For Reading and Listening and Integrated Skills, however, only Part A is compulsory. All candidates must do Part A and choose either Part B1 (easier section) or Part B2 (more difficult section).

To assess the performance of candidates choosing different sections on the same scale, the marks of the easier section are converted to a scale comparable to the one used for the more difficult section. This is done using the marks of the compulsory section as the mediator. The resulting marks are known as the marks of equivalent-difficult-part. A candidate's total paper mark is then calculated by adding the marks of the compulsory part to either the marks of the more difficult part or the equivalent-difficult-part marks. Grading can then be done based on total paper marks for all candidates.

Appendix 3: Latent trait model

A latent trait model developed from modern item response theories is a statistical tool for the calibration of examination items and the estimation of candidates' ability levels based on their responses to items in a test.

The model that is used to calibrate the HKDSE examination questions is based on the Rasch model. The simple Rasch model for dichotomous (i.e. either right or wrong) responses is:

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

However, most examinations contain questions that are not simply right or wrong, and for which more than one mark may be awarded. The model for such polytomous responses is:

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

where β_n is the ability of the person n , δ_i is the difficulty of the item i , $\{\tau_{ik}\}$ are the centralised thresholds and $\sum_{k=0}^0 \tau_{ik} = 0$. β_n and δ_i are generally called the locations (logits).

This is the model that is used to calibrate the items of live examinations together with the data of the research test of a subject. To equate the examination scores across different years, the following formula is adopted to calculate the total score of a candidate on any subset of items:

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

where K_i is the pre-specified multiplying constant of item i , I is the subset of items in the subject, and $E(X_i)$ is the expected value for the candidate n with ability β on item i with the maximum score of m_i :

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

Calibration of all items on different examination papers and the secure research tests enables estimates of the ability of each candidate to be generated on the same scale.

After calibrating item difficulties, it is possible to compile the expected score of a whole subject (which

may consist of different components) for a given candidate's ability by considering the probability of that candidate obtaining a certain mark and different weightings for the different components. These weightings are calculated based on the structure of the examination papers and the performance of the candidates in the examination.

This model supports the grading work of panel judges by generating suggested cut scores for each level of the four core subjects for their reference.

Starting from 2013, in order to provide a basic reference data for equating successive examinations and thus enable standards to be maintained from one year to the next, the HKEAA makes use of the above latent trait model to create an overall calibration of responses to all questions on each examination paper (over two particular years) and all questions on the secure research tests.

Enquiries can be directed to:

Hong Kong Examinations and Assessment Authority

Address : 13/F, Southorn Centre, 130 Hennessy Road, Wan Chai, Hong Kong

Tel : (852) 3628 8860

Fax : (852) 3628 8928

Email : dse@hkeaa.edu.hk