

IB assessment principles and practices: A guide to assessment for teachers and coordinators

Introduction and overview

This publication is intended as a summary guide for teachers and coordinators of the quality methods the IB applies to ensure the assessments that candidates undertake are meaningful and fair.

It covers:

- keeping marking consistent
- improving the quality of marking
- enquiry upon results (EUR) principles.

More detailed information about IB assessment can be obtained from the IB website, www.ibo.org.

What are the stages in the life of an assessment?

The IB utilizes a variety of processes and quality checks to ensure the marks and grades awarded to candidates across all subjects and components are reliable, fair and correct. An assessment is deemed *reliable* if a student would gain the same result if they repeated the assessment on different occasions, and if the assessment was marked by different examiners.

At every stage of the assessment cycle, the IB makes a considerable effort to ensure that assessments are measuring what they should (that is, that they are valid) and that accurate marking standards are applied.

The assessment cycle is a continuous process, whereby each stage is informed by the previous stage and leads into the next stage.

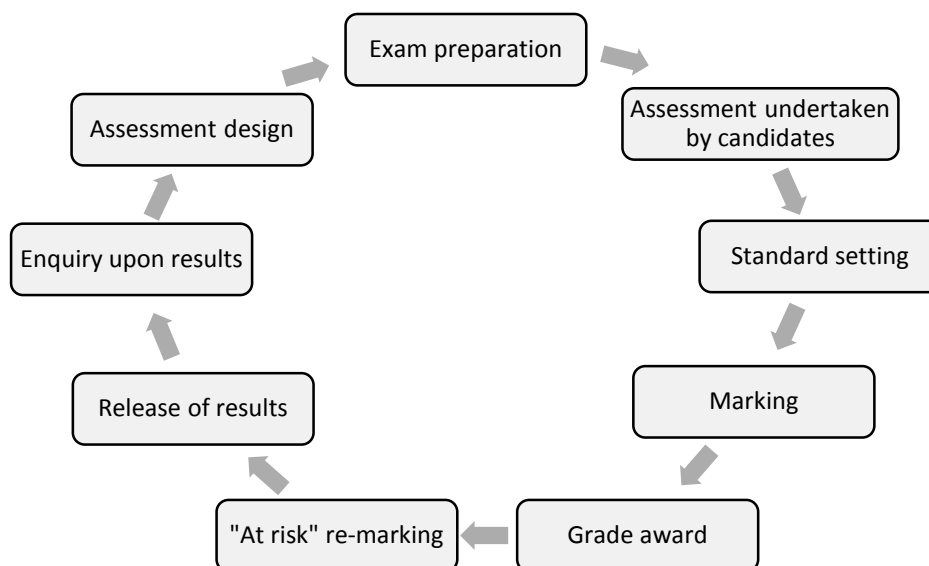


Figure 1. Assessment cycle

How can we be sure all students are treated fairly in the marking?

Students may be assessed by their teacher or by an external examiner. The IB needs to make sure that, in both cases, the same standards are applied.

Research has shown that different examiners will have slightly different opinions on what mark a piece of work deserves. To prevent students being marked differently, the IB appoints an experienced subject expert to be the principal examiner (PE). The PE decides how marks should be awarded, and everyone else has to mark in line with this standard, so the candidate's result does not depend on who marked his or her work.

What is moderation?

Moderation is used with internally assessed work to ensure a common standard across all schools. As a result of moderation, a school's marks may be lowered, raised or remain the same.

The aim of moderation is to check how accurately and consistently the teacher has applied the assessment criteria in his or her marking of the students' work.

The internal assessment (IA) sample is carefully selected to ensure that the mark range of the school is appropriately represented. The IB tends not to select candidates for the moderation sample who have attained full marks, to allow candidates in the higher mark range the possibility of being moderated upwards.

The sample is moderated by the examiner and, based on a statistical comparison between the two sets of marks, an adjustment is made to the teacher's marks for all candidates at the school for that component. If the teacher is consistently under- or over-marking, this adjustment will be the same for each of the teacher's marks. If the teacher is under- or over-marking either at the top or bottom of the mark range, the adjustment may vary across the range of the marks.

In some cases it may not be possible to calculate a moderation adjustment using the submitted sample work. Further work will be requested to ensure that a fair moderation adjustment can be applied. For this reason, all candidates' work must be available until the issue of results.

The moderation process is also applied to a small number of external components. In these cases each examiner sends a sample of their marking to a more senior examiner. Each examiner's marks are then adjusted to bring them into line with the principal examiner's standard of marking.

The moderation hierarchy

Moderation is a hierarchical process designed to ensure that the final marks awarded to every school are in line with the standard of marking that is set by the principal examiner. The different levels of the hierarchy for a typical large-entry IA component where examiners are also subject to moderation are illustrated below.

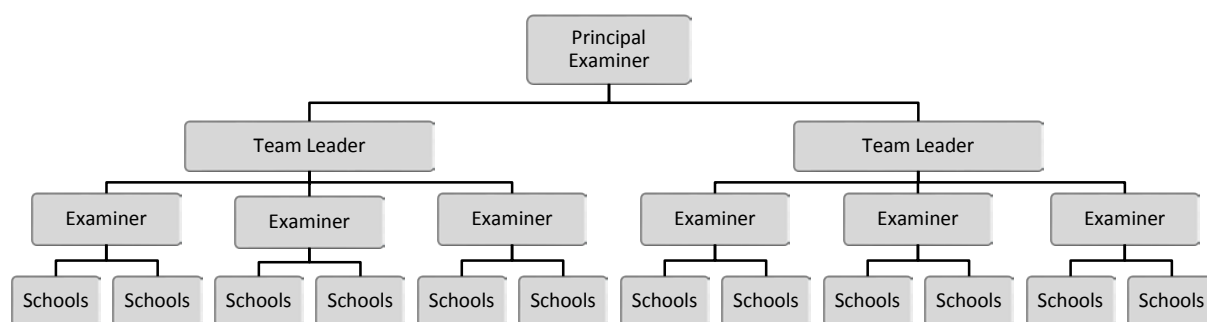


Figure 2. Moderation hierarchy.

The above diagram shows that:

- a school's marks may be adjusted based on the sample submitted to an examiner
- every examiner's marking is also reviewed and adjusted based on a sample of their marking that is submitted to a senior examiner ("team leader")
- in turn, a team leader's marks may be adjusted based on a sample of their marking that is submitted to the principal examiner.

This chain of moderation allows a series of adjustments to be made to a school's marks before a final, moderated mark (aligned to the principal examiner's standard) is awarded. For moderated externally assessed components, the total mark awarded by an examiner on a script is not necessarily the final "moderated" mark awarded to the candidate. The final mark is the result we would expect the principal examiner to award had they marked the work.

How can we be confident that examiners mark students' work correctly?

Before any marking takes place, a standardization meeting is held where the principal examiner, with help from senior examiners, sets the standard of marking for the assessment. All examiners are required to adhere to this standard in their own marking. The standardization team reviews scripts and determines a range of suitable practice scripts, qualification scripts and seeds to help ensure marking reliability of examiners throughout the exam session.

The markscheme is finalized as part of the standardization process in order to make the marking procedure clearer, help the examiners meet the required standard and aid consistency of marking between examiners.

E-marking and seeding

Once exams are taken by candidates, their responses are either sent to scanning centres or uploaded. They are then made accessible to examiners through an online marking tool that randomly allocates responses. The responses are anonymous, which prevents examiner bias.

The process for examiners has three stages: practice, qualification and live marking. Live marking is monitored using seeds.

The practice stage allows examiners to look at scripts that were already marked by the principal examiner (definitively marked) to learn how to apply the markscheme. They receive automatic feedback to clarify their understanding of how marking should be carried out.

The qualification process requires examiners to demonstrate they can apply the markscheme appropriately by testing them with a selection of scripts selected by the principal examiner. An examiner who cannot show that he or she can apply the mark scheme appropriately will not be allowed to start marking candidate work.

Once an examiner has qualified to start marking, his or her marking is monitored throughout by seeds being introduced randomly into their allocation. A seed is a script that has already been definitively marked by the principal examiner. Examiners are unaware that they are marking a seed, which ensures the seeds are marked in the same manner as the other scripts in their allocation.

To help determine that the examiners are marking within the standard set by the principal and the standardization team, each subject is assigned a tolerance value; an examiner's marking should be within this tolerance. A tolerance reflects the legitimate differences in the marks awarded by different examiners to the same piece of work. Think about two teachers in your school marking a piece of work; both agree it is good, but one would award 29 and the other 30.

Marking by item

Marking by item is the same marking process as above. However, it makes scripts easier to mark by splitting the responses into sections known as question item groups (QIG). A QIG can be a question, section or topic. Examiners mark these smaller sections as they do with whole exam responses (that is, completing practice and qualification before going on to live scripts with seeds). Examiners are monitored separately in each QIG, and every QIG has its own marking tolerance.

By splitting exam responses into QIGs examiners can mark those topics in which they are specialists and focus on marking one question at a time, thereby increasing the reliability of marking.

Scaling

In order to aggregate the marks and grade boundaries from the different components to form a percentage total mark, the marks may first need to be scaled. Having scaled marks means that each component has a different weighting based on the component's contribution to the subject's overall assessment.

For example: in Diploma Programme geography standard level, IA has a maximum mark of 30 but is weighted at 25% of the overall assessment. The IA marks are therefore scaled by 25/30, that is, they are scaled down to be out of 25 rather than 30. The paper 1 component has a maximum mark of 60 and is weighted at 40%, which means the marks are scaled down by 40/60. The paper 2 component has a maximum mark of 40 and is weighted at 35%, which means the marks are scaled down by 35/40. The total mark available for the subject is therefore $30 + 60 + 40 = 130$, but once the scaling is applied the maximum "scaled" mark will be 100.

Component	Maximum mark	Weighting	Scaling factor
IA	30	25%	25/30
Paper 1	60	40%	40/60
Paper 2	40	35%	35/40

The overall scaled mark is then used to determine the final grade awarded to the candidate, using the subject grade boundaries.

How are grade boundaries determined?

There is a difference between marking candidates' work and grading their work.

In marking, candidates are given credit for the work they have produced against a markscheme or similar framework. It represents the degree of the assessment task they got right. The mark itself has no other meaning.

In deciding a grade, a judgment is made on the quality of the candidate's work against a defined standard, taking into account the difficulty of the task as well as the proportion of the task that was completed. The grade, therefore, has meaning and relevance and is intended to be comparable with performances on other assessments.

Consider: *Marks* are how far a candidate has walked, but *grades* take into account how steep the path was.

Grade awarding is the means by which the IB converts marks into grades and takes place just before the issue of results. Senior examiners, including the principal examiner, and IB staff set the grade boundaries and resolve any outstanding issues from the exams. The grade boundaries are determined using a combination of judgment, statistical evidence, grade descriptors and the G2 forms submitted by teachers. G2 forms provide important information for senior examiners, helping them to agree appropriate grade boundaries in light of teachers' perceptions about the examinations.

How can we be confident the results are correct?

The use of qualification scripts and seeds means the IB is confident that each script was marked correctly. However, because it is important for candidates to receive the correct subject grade based on the work they submitted, additional quality checks on the marking of candidate work are made.

“At risk” re-marking

The intention of the “at risk” process is to identify and re-mark candidates whose final grades are at risk of being too low based on their predicted grades. It also serves to detect any marking problems that may not have been identified previously. This practice ensures that borderline candidates receive an accurate grade.

Enquiry upon results

If a candidate or school believes that the grade awarded is incorrect for some reason, they can request an enquiry upon results (EUR). After the issue of results, the coordinator may request a:

- category 1: re-mark of all of a candidate’s externally assessed components for a subject
- category 2: copies of externally assessed component material
- category 3: re-moderation of a school’s sample of an internally-assessed component
- category 1 report: report on the marking of a category 1 EUR.

A fee is payable for each of the above categories, except when a grade is changed as a consequence of category 1 re-mark.

A returned script may contain useful comments from the allocated examiner. However, this can’t be guaranteed because examiners are not required to write comments when marking candidates’ work. See the next section for more information about receiving feedback.

As mentioned in the “What is moderation?” section, the mark awarded by an examiner as shown on a script may not necessarily be the final “moderated” mark awarded; the mark awarded to the candidate in IBIS is correct.

What feedback is available?

Approximately one week after the issue of results, all available IA feedback forms are released on IBIS. The IA feedback form is an opportunity for examiners to comment on suitability of tasks and any administrative issues. Examiners who are subject to moderation are advised not to comment specifically about the teacher’s marks of the IA sample because they are not aware if a further (examiner) moderation factor may also be applied. Examiners who are not subject to moderation are able to make more detailed comments about the adjustments they have made to the teacher’s marks.

Examiners marking externally assessed components are not required to write comments when marking candidates’ work because the emphasis for examiners is to make sure the work is marked correctly, rather than to provide notes and recommendations on the work itself. Feedback on candidate performance is an important part of formative assessment that is carried out by the teacher throughout the course.

The online curriculum centre (OCC) is another source of feedback where both subject reports and teacher support material can be found.

How are exams created?

In order to meet the challenge of producing valid exams, work may begin on preparing a set of exams up to two years before it is taken by candidates. As part of this process, an exam editing meeting is held that involves the subject manager, component principal examiners and a selection of other examiners. They

discuss the exams and markschemes before the assessments are finalized, ensuring that the assessment standard is at the correct level. The set of exams for each level is reviewed as a whole, to see how effectively course content and course objectives are covered. The group also critically reviews each question, looking at its suitability, accuracy and any potential for ambiguity or bias. They will ensure that markschemes are clear, concise and comprehensive for the examiners to follow.

What is done to improve the quality of assessment?

As the assessment process is a continuous cycle, the experience of previous sessions informs developments in the assessment design and exam preparation for future exam sessions. The IB will continue to progress in the development of quality control mechanisms to ensure that assessments result in valid and reliable outcomes for candidates. More information about IB Assessment can be found on the OCC and on the IB website.