International Baccalaureate®
Baccalauréat International
Bachillerato Internacional

Assessment principles and practices: Quality assessment in a digital age

International Baccalaureate®
Baccalauréat International
Bachillerato Internacional

Assessment principles and practices: Quality assessment in a digital age

Assessment principles and practices: Quality assessment in a digital age

Published May 2026

Published by the International Baccalaureate Organization, a not-for-profit educational foundation of Rue du Pré-de-la-Bichette 1, 1202 Genève, Switzerland.
Website: ibo.org

© International Baccalaureate Organization 2026

The International Baccalaureate Organization (known as the IB) offers four high-quality and challenging educational programmes for a worldwide community of schools, aiming to create a better, more peaceful world. This publication is one of a range of materials produced to support these programmes.

The IB may use a variety of sources in its work and check information to verify accuracy and authenticity, particularly when using community-based knowledge sources such as Wikipedia. The IB respects the principles of intellectual property and makes strenuous efforts to identify and obtain permission before publication from rights holders of all copyright material used. The IB is grateful for permissions received for material used in this publication and will be pleased to correct any errors or omissions at the earliest opportunity.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the IB's prior written permission, or as expressly permitted by the Rules for use of IB Intellectual Property.

IB merchandise and publications can be purchased through the IB Store (email: sales@ibo.org). Any commercial use of IB publications (whether fee-covered or commercial) by third parties acting in the IB's ecosystem without a formal relationship with the IB (including but not limited to tutoring organizations, professional development providers, educational publishers and operators of curriculum mapping or teacher resource digital platforms, etc.) is prohibited and requires a subsequent written licence from the IB. Licence requests should be sent to copyright@ibo.org. More information can be obtained on the IB public website.

International Baccalaureate, **Baccalauréat International**, **Bachillerato Internacional** and IB logos are registered trademarks of the International Baccalaureate Organization.

IB mission statement

The International Baccalaureate aims to develop inquiring, knowledgeable and caring young people who help to create a better and more peaceful world through intercultural understanding and respect.

To this end the organization works with schools, governments and international organizations to develop challenging programmes of international education and rigorous assessment.

These programmes encourage students across the world to become active, compassionate and lifelong learners who understand that other people, with their differences, can also be right.

IB learner profile

The aim of all IB programmes is to develop internationally minded people who, recognizing their common humanity and shared guardianship of the planet, help to create a better and more peaceful world.

As IB learners we strive to be:

INQUIRERS
We nurture our curiosity, developing skills for inquiry and research. We know how to learn independently and with others. We learn with enthusiasm and sustain our love of learning throughout life.

KNOWLEDGEABLE
We develop and use conceptual understanding, exploring knowledge across a range of disciplines. We engage with issues and ideas that have local and global significance.

THINKERS
We use critical and creative thinking skills to analyse and take responsible action on complex problems. We exercise initiative in making reasoned, ethical decisions.

COMMUNICATORS
We express ourselves confidently and creatively in more than one language and in many ways. We collaborate effectively, listening carefully to the perspectives of other individuals and groups.

PRINCIPLED
We act with integrity and honesty, with a strong sense of fairness and justice, and with respect for the dignity and rights of people everywhere. We take responsibility for our actions and their consequences.

OPEN-MINDED
We critically appreciate our own cultures and personal histories, as well as the values and traditions of others. We seek and evaluate a range of points of view, and we are willing to grow from the experience.

CARING
We show empathy, compassion and respect. We have a commitment to service, and we act to make a positive difference in the lives of others and in the world around us.

RISK-TAKERS
We approach uncertainty with forethought and determination; we work independently and cooperatively to explore new ideas and innovative strategies. We are resourceful and resilient in the face of challenges and change.

BALANCED
We understand the importance of balancing different aspects of our lives—intellectual, physical, and emotional—to achieve well-being for ourselves and others. We recognize our interdependence with other people and with the world in which we live.

REFLECTIVE
We thoughtfully consider the world and our own ideas and experience. We work to understand our strengths and weaknesses in order to support our learning and personal development.

The IB learner profile represents 10 attributes valued by IB World Schools. We believe these attributes, and others like them, can help individuals and groups become responsible members of local, national and global communities.

International Baccalaureate®
Baccalauréat International
Bachillerato Internacional

© International Baccalaureate Organization 2017
International Baccalaureate® | Baccalauréat International® | Bachillerato Internacional®

Contents

Introduction

The International Baccalaureate (IB) programmes are offered in over 150 countries and regions (territories), encompassing a diverse range of educational contexts and traditions. For some, the IB's assessment philosophy and methods will be familiar; for others, the system might seem complex and unfamiliar.

The IB undertakes a challenging task through its assessment principles by committing to assess what is truly important rather than what is simply easy to measure. Alec Peterson, the first Director General of the IB, articulated this challenge within the educational ecosystem as requiring balance.

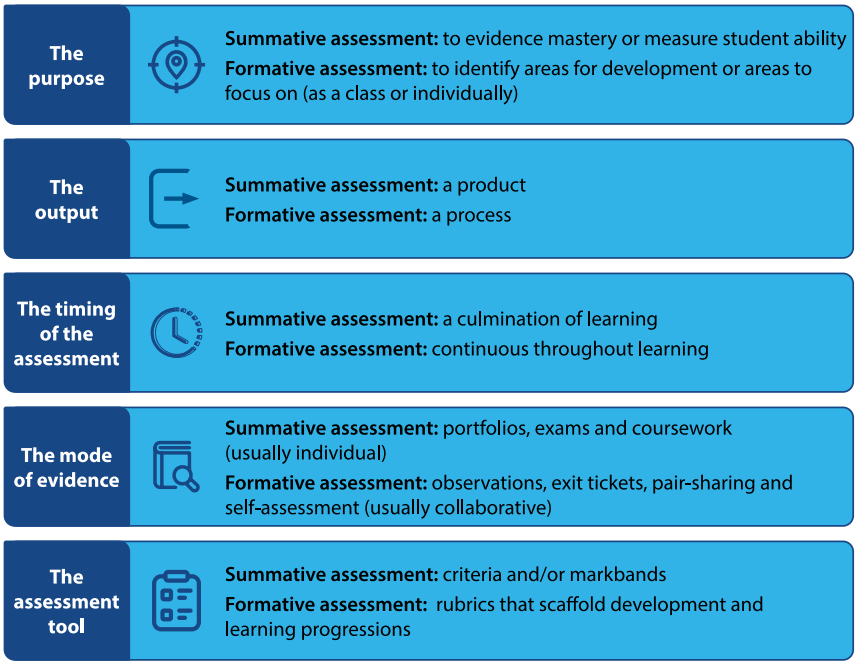
> What is needed is a process of assessment which is as valid as possible, in the sense that it really assesses the whole endowment and personality of the pupil in relation to the next stage of [their] life, but at the same time sufficiently reliable to assure pupils, parents, teachers and receiving institutions that justice is being done. Yet such a process must not, by its backwash effect, distort good teaching, nor be too slow, nor absorb too much of our scarce educational resources.
>
> (Peterson, 1971, pp. 27–55)

This publication aims to demystify the principles that the IB employs to ensure that assessment across its programmes is meaningful, fair and beneficial for students in a manner that is authentic, relevant and reflective of the learning and teaching process.

Assessment in the IB is broadly categorized into two types, as shown in figure 1.

Figure 1

Summative and formative assessment



Understanding the distinct roles of these assessments is crucial for the IB community.

Assessment results significantly impact students' lives, influencing their progression into further study or careers. Assessment is a tool that can lead to both positive and negative outcomes. As teachers, parents and legal guardians, or students, understanding the strengths, weaknesses and decisions behind assessments is crucial. This publication addresses common questions and concerns such as:

- "Why would the IB do that?"—understanding the principles behind IB assessment practices

- "It's not fair. I needed a higher grade"—explaining the fairness and implications of assessment decisions
- "I hadn't thought about it like that before"—unpacking the assessment process
- "I only want to use assessment to inform better teaching"—highlighting the importance of high-quality assessments for both formative and summative purposes.

Technological shifts are happening constantly, and embracing technology through IB assessments is critical in order that assessment remains authentic and reflective of the ways in which students engage, are taught and how they learn. Including "quality assessment in a digital age" in the title of this resource recognizes that IB assessments need to be delivered in a way that students are familiar and comfortable with, if they are to remain relevant. More importantly, it recognizes that, when making the transition, the IB needs to remain faithful to the underlying principles on which it is built. This means that the technology is used to support learning and assessment in order to create more valid assessments. The description "quality assessment in a digital age" underlines the IB's enthusiasm for, and responsiveness to, this process and how the landscape is evolving, while reinforcing its commitment to fundamental IB assessment principles.

This publication provides a detailed explanation of the IB's assessment principles as they apply to all its programmes: the Primary Years Programme (PYP), the Middle Years Programme (MYP), the Diploma Programme (DP) and the Career-related Programme (CP). These principles cover all types of IB assessments and explain why they are designed the way they are. While some sections focus specifically on summative assessment, including assessment submitted to the IB for external marking, the insights provided are intended to enhance educational practice across all programmes. Even in contexts where external summative assessments are not conducted, understanding the elements of effective and fair assessment can improve classroom experiences and better prepare students for future testing. Further, the publication intends to unpack assessment principles for the entire IB community, including teachers, school administrators, examiners, and university and college admissions professionals. It is written for those who are familiar with the IB and its assessments, but who may not have engaged deeply with the underlying theories of assessment. It is hoped that this resource helps explain how IB assessment principles support the wider educational intentions of the IB in providing a world-class experience for its students.

Terminology

IB assessments use a large number of terms that may be unfamiliar to teachers and students. Throughout all the IB's publications, an effort is made to avoid confusion while remaining precise.

As part of this commitment, abbreviations are spelled out in full the first time they are used in each section of this publication. Similarly, important terms are explained the first time they are used, either directly or through linked material.

This publication also has a full list of assessment terms in the glossary. Outside of general assessment terms, subject-specific terminology is explained in the respective IB subject guide.

Additional resources

This resource is one of several published by the IB to explain its approach to education. It can often be confusing to understand how these key resources relate to each other. Tables 1 and 2 provide a summary of key resources and an explanation of their purpose and role.

Table 1

Key IB publications for cross-programme use

Key IB publications—cross-programme	
Assessment principles and practices: Quality assessment in a digital age	This publication explains the overarching approach that the IB takes to assessment and how it intends to apply this in practice. It focuses on the summative assessment (formal testing) element of an IB education.
What is an IB education?	The aim of this publication is to communicate clearly what lies at the heart of an IB education. It explains the ideals that underpin all IB programmes. By describing the IB's educational philosophy, the publication also offers support for schools on their journey from adopting the IB through many years of being an IB World School.
Programme standards and practices	This publication provides a set of criteria against which both schools and the IB can evaluate success in the implementation of the four IB programmes. It contains programme standards (general requisites for schools to implement any IB programme), practices (further definitions of the standards, which are common to all programmes) and requirements (specific to each individual programme).

Table 2

Key IB publications for specific programmes

Key IB publications—programme-specific	
Primary Years Programme: From principles into practice *Middle Years Programme: From principles into practice* *Diploma Programme: From principles into practice* *Career-related Programme: From principles into practice*	These publications focus on learning and teaching in the context of a particular IB programme. They also explain the requirements of the programme.
Subject guides (MYP and DP only)	These contain detailed information on one subject, course or discipline, for example, geography or visual arts. They set out the aims, objectives, syllabus and criteria for internal assessment (IA) in a particular subject. They usually also provide

Key IB publications—programme-specific	
	additional subject-specific guidance for learning and teaching.
Subject continuums (PYP only)	The subject continuums are designed as examples that schools can use to support the design of their own scope and sequences. They can be adapted as needed to fit local community contexts and requirements. The continuums also provide conceptual understandings and example concepts for each subject and its strands. The accompanying subject overviews make transparent the essential elements of the PYP in the context of the subject areas and clarify the role of the subject areas in a transdisciplinary programme.
Middle Years Programme Assessment procedures *Diploma Programme Assessment procedures* *Career-related Programme Assessment procedures*	These publications set out the rules, regulations and specific processes that must be followed in delivering a particular IB programme's assessment. They contain general regulations: the regulations that underpin the rules that IB World Schools are expected to follow.
Preparation for examinations policy (MYP, DP and CP only)	This publication outlines the IB's policy on the preparation for examinations, including the secure storage of IB confidential materials on school premises and the rescheduling of examinations.
Conduct of examinations booklet (DP only; CP uses the DP version) and *The conduct of IB Middle Years Programme on-screen examinations* (MYP)	This publication informs coordinators and invigilators of the regulations concerning the administration and conduct of a programme's examinations. A copy of this publication must be available in every examination room.

In addition to these high-level publications, the IB provides a range of other resources that offer focused guidance to teachers, coordinators and other stakeholders. These can be found on the IB website and the Programme Resource Centre (PRC).

Using this resource

The range of topics covered in *Assessment principles and practices: Quality assessment in a digital age* is broad. The publication has been designed to have many routes through it, depending on the interests and needs of the reader. The **topic questions** in this section can act as starting points for learning about those aspects that are most important to a given reader.

Topic questions

Assessment principles and practices: Quality assessment in a digital age is intended as a comprehensive overview of the way that the IB approaches assessment; however, many readers will be looking for specific information about part of the assessment process. To that end, the topic questions are followed by lists of sections that concentrate on particular questions readers may have. These lists are intended as references only, and not as a suggested order in which the sections should be read.

The glossary at the end of this publication can also be used to identify terms of interest. Use the "Find" option in the PDF version of this publication to find out more about any terms of interest in the glossary.

"What is the IB's approach to assessment?"

For university admissions staff, teachers and stakeholders who are interested in the theoretical underpinnings of the IB assessments, how the IB defines good-quality assessments, what directs its decision-making in setting processes, and how it balances the conflicting demands of assessment.

Typical questions include the following.

- What is IB assessment all about? What makes it special?
- How is artificial intelligence (AI) changing assessment?

Sections that are particularly relevant for these questions include the following.

Introduction and overview

- Good assessment
- Good assessment supports curricular goals
- Good predictability
- Good assessment uses a range of assessment tasks
- The role of classroom-based assessment and internal assessment
- Collaborative working and individual marks
- Good assessment considers wider student competencies and higher-order thinking skills
- Higher-order thinking skills
- The learner profile and skill development
- International-mindedness and intercultural understanding
- What does good digital assessment look like?

Section A: Principles of assessment

- Assessment in education
- Defining assessment
- Approaches to assessment
- Backwash effect and learning

- Validity
- Defining validity
- Creating a validity argument
- Maintaining validity
- Elements of the validity chain
- Balancing aspects of validity
- Reliability
- Consistent outcomes and the "right" outcome
- Construct relevance and authenticity
- Manageability
- Fairness and bias
- Comparability
- The IB's approach to validity
- Benefits of digital assessment to validity
- Standards
- Three aspects of standards
- Norm-referencing and criterion-referencing
- Norm-referencing
- Criterion-referencing
- Attainment-referencing
- Maintaining standards
- Describing student success in summative assessment
- The impact of grades
- The importance of professional judgement
- Marking IB assessments
- Defining marking
- Marking methods
- Marking formative assessment

Section B: IB assessment practices

- Reporting student achievement
- The meaning of IB grades
- The difference between marks and grades
- A successful examination session
- Grades and achievement
- Defining academic integrity
- Prioritizing fairness for all

Printable resources

- "The IB's principles of assessment" (PDF)

"I am a PYP practitioner—we don't do assessment. How does this publication benefit me?"

For PYP practitioners who want to understand school-based summative assessment and formative assessment opportunities.

While there are no assessments submitted to the IB for the PYP, school-based summative and formative assessment opportunities do exist in the PYP, and this publication outlines good practice for those areas. Policy areas around academic integrity and nurturing ethical mindsets are also important to discuss during early years, in ways that are meaningful for the age of the students.

Sections that are particularly relevant for these questions include the following.

Section A: Principles of assessment

- Approaches to assessment
- Marking formative assessment
- Nurturing an ethical mindset
- Ethical use of artificial intelligence in assessment

Section B: IB assessment practices

- Defining academic integrity
- Prioritizing fairness for all

Section C: IB programme-specific processes

- Elements common to all programmes
- Primary Years Programme

"I am an MYP practitioner—we are already assessing digitally. How does this publication benefit me?"

While there are elements of this publication that discuss the transition to digital assessment in the Diploma Programme and Career-related Programme (DPCP), these sections also attest to the actual value of MYP eAssessments. The publication applies to the MYP. Please see the topic question "I am an MYP, DP or CP practitioner interested in how IB-submitted assessments are marked and grades awarded. What feedback is available for schools?".

"What difference will digital assessment make to DPCP assessment?"

For teachers who are familiar with IB assessment and are interested in how the IB will handle the move to digital assessments.

Typical questions include the following.

- What is digital assessment?
- Will work be marked by a computing device?
- Why will it be better?
- Will it be the same standard?

Sections that are particularly relevant for these questions include the following.

Introduction and overview

- Good assessment
- The role of technology in IB assessment
- What does good digital assessment look like?
- Risks associated with digital assessment

Section B: IB assessment practices

- Prioritizing fairness for all
- The assessment life cycle
- The impact of digital assessment on the assessment cycle

"I am an MYP, DP or CP practitioner interested in how IB-submitted assessments are marked and grades awarded. What feedback is available for schools?"

For teachers who want to understand what goes on once they send students' work to the IB for marking, and how grades are produced. These teachers may also want to understand why the IB includes teacher assessment in its processes but does not always accept teachers' judgements.

Typical questions include the following.

- Why are there different grade boundaries this year?
- Why can't I have a different examiner?
- Who marks the scripts?
- How are the examiners checked?
- How can I appeal?
- I would have given a different mark to the student—why is this fair?
- The markscheme indicates that the student's answer is worth a mark—how can this be fair?
- Isn't a "grade-free classroom" the best approach for students?
- Why are my marks changed?
- Why does my moderation factor change between years?

Sections that are particularly relevant for these questions include the following.

Introduction and overview

- The role of classroom-based assessment and internal assessment

Section A: Principles of assessment

- Assessment in education
- Defining assessment
- Approaches to assessment
- Backwash effect and learning
- Elements of the validity chain
- The impact of grades
- Marking IB assessments

Section B: IB assessment practices

- Reporting student achievement
- The meaning of IB grades
- The difference between marks and grades
- A successful examination session
- Grades and achievement
- Predicted grades
- Assessment process: Roles and responsibilities
- Principal examiner and chief examiner
- Other examiner roles
- The responsibility of IB staff
- Examiner hierarchy
- Prioritizing fairness for all
- Marks and grades
- Approaches to marking
- Analytic markschemes

- Holistic criteria: Markbands
- Standardization
- The quality model
- Practice
- Qualification
- Live marking and seeds
- Indicators of successful standardization
- Tolerances
- Challenging and unusual scripts
- School connections
- Question item groups
- Aggregation
- Moderation
- Moderation and expectations of teachers
- Selection of student work
- Irregular situations
- Failing to find a moderation factor
- Dynamic sampling
- Grade awarding and aggregation
- Judgemental and interpolated grade boundaries
- Evidence used in the grade award process
- Considering this year's cohort
- Feedback on the assessment
- Reviewing script evidence
- Reviewing statistics on outcomes
- Balancing the evidence
- Fixed grade boundaries
- Review of the grade boundary recommendation and approval
- Awarding a programme certificate
- Teacher observers
- Principles of the grade award process
- Quality checks
- Feedback to schools
- Internal assessment feedback

Section C: IB programme-specific processes

- Elements common to all programmes

"How does the IB design and develop examinations?"

For teachers who want to understand how examinations are crafted, and for those who have concerns about the quality of them.

Typical questions include the following.

- Who writes examinations?
- How are they checked?
- Are they the same when in different languages?

- Do they really test the right thing?

Sections that are particularly relevant for these questions include the following.

Introduction and overview

- Good assessment
- Good assessment supports curricular goals
- Good predictability
- Good assessment uses a range of assessment tasks
- The role of classroom-based assessment and internal assessment

Section A: Principles of assessment

- Assessment in education
- Defining assessment
- Approaches to assessment
- Backwash effect and learning
- Validity
- Defining validity

Section B: IB assessment practices

- Roles in authoring examinations
- Prioritizing fairness for all
- Assessment production
- Overview of examination production
- Process up to content sign-off
- Authoring and content sign-off
- Building and proofreading
- Usability sign-off
- Quality control
- Translations
- Modified papers
- Marks and grades

Assessment principles and practices

Assessment principles are what is judged to be important in creating, delivering, marking and grading qualifications and assessments. They come from what we think is important about an IB education. The most important principle is that assessments should support education, and not distort it.

Assessment practices are the ways in which the IB delivers its principles meaningfully and practically. They take into account the conflicting demands and practical limitations of working in the world while maintaining the IB philosophy of being principled.

The following points summarize the IB's principles of assessment.

Assessment, whether summative or formative, must:

- be valid for the purpose for which it is intended. This means it must be balanced between the conflicting demands of "construct relevance" (that is, testing the right thing), reliability, fairness (that is, no bias), comparability with alternatives, and manageability for students, schools and the IB
- have a positive backwash effect—that is, the design must encourage good-quality learning and teaching
- be appropriate to the widest possible range of students, allowing them to demonstrate their personal achievement level
- be designed so that all necessary arrangements can be made to ensure access for all students
- be part of the context of a wider IB programme, not considered in isolation
- support concurrency of learning and the overall student experience
- support the IB's wider mission and develop learners who are inquirers, knowledgeable, thinkers, communicators and open-minded.

Assessment practices are generally described at a high level and should be implemented appropriately within the context of individual subjects, disciplines and programmes. Assessment practices explain broadly what needs to be done, but without detail on the day-to-day process and the differences in implementation across individual programmes and subjects.

The printable resource "The IB's principles of assessment" (PDF) is available for teachers.

Good assessment

- There is no single definition of "good assessment"—it depends on the relative importance placed on different priorities, and the purpose of the given assessment.
- At the IB, the underlying principle is to test what is important rather than to judge as important what can be tested.
- IB assessment seeks to have a positive backwash effect on learning and teaching.
- In general, good assessments should include a range of performance tasks that help evidence a well-rounded profile of student learning, and include the opportunity for more in-depth classroom-based activities, as well as examinations.
- To include all students with diverse learning styles and preferences, good assessments should be designed and developed with universal design for assessment principles in mind. They should consider multiple means of engagement, representation, action and expression.

A **good assessment**, in the IB's view, **prioritizes meaningful learning outcomes**. It aligns assessment closely with curricular goals and supports students in developing a broad range of skills and competencies.

The IB's underlying principle is to **test what is important**, rather than to judge as important what can be tested. This needs to be balanced against other considerations such as reliability and student workload.

It is difficult for any single approach to be successful in delivering every possible priority. In particular, good assessment design is different for summative and formative assessment. Expanding on this principle, the IB considers a good assessment to be one that:

- supports curricular goals
- uses a range of assessment tasks
- considers wider student competencies and higher-order thinking skills.

The following sections consider each of these in turn.

Good assessment supports curricular goals

- Assessments should have a positive backwash effect, encouraging good learning and teaching.
- Assessments should have good predictability.

Assessment should not be seen as separate to learning and teaching. IB assessment outcomes are based on summative assessment and are not intended to provide direct feedback on learning and teaching. However, it is well understood that what is included in the assessment will have an impact on what is taught. This is known as the "backwash effect".

For the IB, the design of assessments should encourage the most desirable educational outcomes for all students. The impact on student learning remains an essential consideration in the design of IB assessments and, together with construct relevance, this consideration helps guide how the IB balances the different elements of validity.

The impact of high-stakes assessment on learning and teaching can be used advantageously by designing assessments that encourage good pedagogy and constructive student involvement in their own learning (see, for example, Murphy, 1999).

Throughout IB programmes, assessment is closely tied to, and reflective of, the learning experience. Using the constructive alignment approach (Biggs, 1996), everything is designed around the intended learning

outcomes. These outcomes shape and steer both how students are taught and how they are assessed, ensuring they are well prepared to succeed. For more information, see *Approaches to assessment in the IB*.

The desired personal characteristics of students, expressed in the IB mission statement and explored in the learner profile, fit very well with a **constructivist theory of student learning**. In this framework, students actively engage in the learning process, take responsibility for their own learning, and grow their knowledge, understanding and skills through inquiry.

For example, active and open engagement with cultural perspectives other than the student's own is an explicit expectation in the assessment requirements of a number of subjects. The more affective qualities of care and compassion are difficult to include in formal assessment, but nevertheless must be represented within the overall assessment system. This is partially achieved through programme-specific requirements of the curriculum that are not assessed, such as the exhibition in the PYP, the community engagement element of the MYP, the creativity, activity, service (CAS) requirement in the DP, and community engagement in the CP.

In terms of the design of DPCP courses particularly, the IB places a strong emphasis on predictive validity (the degree to which the results predict future success), with an awareness that the manner in which assessment is conducted will have a major impact on how IB courses are taught within schools. The assessment model applied to each subject is designed to be broad, to include a variety of types of evidence, and to support construct relevance by giving the broadest range of evidence to support student achievement and learning.

While the IB is very aware that assessments have a backwash effect on learning and teaching, it also encourages schools to adopt pedagogies that develop all the goals and philosophies of IB programmes in the students.

The IB regularly undertakes research studies to evaluate the extent to which it has been successful in designing programmes that are good preparation for further study. For more details on the range of research undertaken by the IB, refer to the "Research" section of the IB website.

Good predictability

"Predictability" describes being able to gauge what will happen or when something will happen. In assessment, predictability means schools' ability to determine what kinds of questions will be asked on a given examination, and when. Good predictability in assessment helps create fairness by aligning questions with the curriculum that teachers use to prepare students. This ensures that students know what general types of questions to expect, even if the exact content varies.

The IB seeks to ensure that schools' investment in their teaching options is rewarded over the whole lifetime of a particular curriculum (before it is reviewed). The entire syllabus should be examined in the way that the specific assessment dictates. Care is taken towards the end of a course to eliminate the inevitability of bad predictability (where a school identifies what has not yet been asked, and is therefore likely to appear in an examination).

The underlying principle is that nothing in a given examination should be a surprise, either for the student or the school. The questions asked, therefore, should be explicitly supported by subject guides for any given component. Where possible and appropriate in assessments, the IB seeks to reduce the likelihood of problems caused by predictability. It does so by designing assessments in such a way that students with pre-prepared answers are not advantaged.

Assessment design is paramount in ensuring that there are sufficient ways to test any given theme, option or text in order to mitigate bad predictability.

Table 3

Good and bad predictability

Good predictability	Bad predictability
If every permutation of a question was mapped based on a theme, option or text plus a command	Teachers overly prepare students on certain questions where they have noticed a tendency for a very similar question to be asked in repeated

Good predictability	Bad predictability
term, the questions asked by the IB for that subject would feature in the results.	examination sessions. This is problematic for longer-response questions that require demonstration of multiple skills (knowledge, analysis and evaluation).
An unpredictable approach to reusing questions.	There are a limited range of questions that can be asked, owing to ineffective assessment design decisions (for example, setting a prescribed text for a course that has only one main theme on which questions can be asked). This leads to bad predictability by design.

Good assessment uses a range of assessment tasks

A multiple-choice question, a short-response question, an extended-response question, an essay, a project, a single piece of work from a portfolio, and a research assignment are all examples of assessment tasks.

An assessment instrument or component is made up of one or more tasks that are collected together on the basis of thematic or content continuity, or for convenience. A summative examination, a portfolio of student work, an extended collaborative project, and a research assignment are examples of assessment instruments or components. There is overlap between the concept of an assessment task and the concept of a component. Sometimes, a student may undertake only one task from a number of choices available for a component.

There are a number of reasons why the IB uses a wide variety of assessment tasks and components. First, from a historical and pragmatic perspective, Peterson (2003) says of the original development of DP assessment that "we had both an obligation and an opportunity to take into account the differing techniques of assessment used in those countries and regions (territories) to whose institutions IB students were mostly seeking entry", and this principle extends to the MYP and CP. There are also validity considerations, relating to fitness for purpose, that require a varied approach to assessment. Finally, a variety of assessment techniques helps to reduce the potential for inequity in assessment—see Linn (1992) and Brown (2002). The range of components and the setting of tasks within them ensure that, taken across the assessment model for a whole subject, student achievement is adequately represented against all the objectives for that subject.

The role of classroom-based assessment and internal assessment

Classroom-based assessment offers a number of opportunities to evaluate students in areas that are not well suited to examinations. The most important aspect of this is that students can be asked to perform an extended task that gives them the opportunity to investigate a problem and show how they develop their thinking without the time pressures that are inherent in an examination. This means that there are a wide range of assessment tasks that can only be delivered using classroom-based assessment.

Examples of the kind of tasks that lend themselves to classroom-based assessment include project work, fieldwork, laboratory practical work, performances and mathematical investigations.

There are other **advantages** to internal assessment (IA) within the context of an international qualification. Such assessment can offer greater **flexibility** in the choice of topic, while continuing to address a common set of skills. This allows schools to place study in a **local, cultural or geographical context**, or to draw closer links between the classroom and the world beyond it. International schools, whose students often have a different cultural background from that in which the school is embedded, can use internally assessed work to develop a closer involvement in the local society or environment.

Additionally, IA can often provide individual students with the opportunity to select their own topic or issue. This allows students to follow a particular interest and gives them greater control over their own learning. This flexibility of approach makes IA a valuable addition to students' education, improving the validity not only of the assessment process, but also of the learning experience as a whole.

There are also **challenges** around classroom-based assessment. Key among these is ensuring **academic integrity**. When assignments are completed without standardized (supervised) conditions—outside of the classroom, for example—it is much harder to ensure that students are not engaging in activities that contravene the IB's academic integrity policy, such as asking someone else to produce their work for them. With open access to the internet, this challenge is compounded exponentially. This is why the IB considers school staff and teachers best placed to help authenticate student coursework and identify where students are not submitting their own work. More information and resources related to academic integrity can be found on the "Academic integrity" page of the IB website.

Another challenge with classroom-based assessment is that it can generate a significant **burden** on both the teacher and student. Undertaking assessment in class reduces the amount of teaching time available, and internally set tasks usually require a significant time commitment from the student. While it is appropriate for teachers to spend time preparing students with the skills and processes required for IA, there may be a strong temptation, felt by both student and teacher, to rehearse and practise the particular task set for IA more than is necessary, further reducing the time spent teaching. DP subject guides offer detailed guidance regarding the role of the teacher in supporting IA work.

Classroom-based assessment can be assessed either **externally** or **internally**. When externally assessed, the work produced by the student is sent to the IB and assessed by an examiner. The quality of examiners' decisions are monitored by the IB (see the section on "Marking" within this publication).

By contrast, for IA, the school (typically the student's teacher) marks the student's work, and the IB checks that the teacher has correctly applied the global standard through a process of moderation (see the section on "Moderation" within this publication).

There are different views on IA around the world. Some education systems place great value on the fact that teachers are best placed to give a holistic opinion on the performance of the student rather than just having a "snapshot" from a single assessment. Other systems rate teacher performance based on their students' outcomes, which creates a strong incentive for teachers to award high marks in any IA.

A teacher's judgement can also be affected by past experience of a student's work, which can introduce various forms of bias. Teachers may sometimes be unclear about the limits of their role in guiding and supporting students as they carry out internally assessed work, and may often have only a limited view of global standards of achievement within their subject area. When assessing their own students' work, teachers may also be heavily influenced by the general standards existing within their own schools. Even where there are no incentives to over-reward work submitted for assessment, the professional relationship that teachers establish with students makes objective decisions challenging. Research on this unconscious bias suggests that it can have a noticeable effect (see, for example, Zanga, De Gioannis, 2023).

The IB believes that the benefits gained from setting meaningful assessment tasks through classroom-based work evaluated by someone who has seen the students' development are greater than the risks that IA creates. Therefore, an IA task will usually be an integral component in a subject's assessment model.

Collaborative working and individual marks

One challenging area is how to assess collaborative tasks, where students are expected to work together to complete an activity. This is an important aspect that students will encounter in their future workplaces, but one that is often neglected in academic assessment.

In some circumstances, it is possible to identify the individual's role in a group activity. An example might be a dance performance, where one can see the skills demonstrated by each individual dancer and mark them individually. In these situations, group work is not contentious.

In other situations, it is not possible to identify who is responsible for what aspect of the final work. In such cases, it is possible to award a single mark for the entire group, but this does not take into account differences in the contribution each student has made to the overall result. One student might have contributed most of the work and expertise, but they would get no more credit than any other student. Given the individual nature of the way that IB results are used to make selection decisions, the IB does not believe this is a fair approach. In general, it tries to avoid group assessments where individual student achievement cannot be measured.

Good assessment considers wider student competencies and higher-order thinking skills

An IB education seeks to achieve more than getting students to learn "facts". This long-standing goal of the IB is reflected by current thinking among governments for the need to provide students with 21st-century **skills**, **workplace competencies** or similar **initiatives**.

The IB's philosophy and approach to student competencies are focused on the **learner profile** and its link to **international-mindedness**. This is expanded upon in "Section C: IB programme-specific processes", which explores continuity between programmes. The most important concept to keep in mind is that the IB's approach to assessment ensures that students are evaluated not just on content knowledge, but on a range of competencies that reflect real-world applications. Through thoughtful design, IB assessments aim to cultivate **lifelong learners** who are well prepared for future challenges.

Therefore, a good assessment considers the **full range of outcomes** that each course seeks to achieve and allows students to demonstrate their abilities in all of these. However, it is often the case that it is only desirable or manageable to measure a fraction of these outcomes, and good-quality assessment also balances these limitations. For the IB, these outcomes can best be categorized by **higher-order thinking skills**, wider student **competencies** and their link to **international-mindedness**.

Higher-order thinking skills

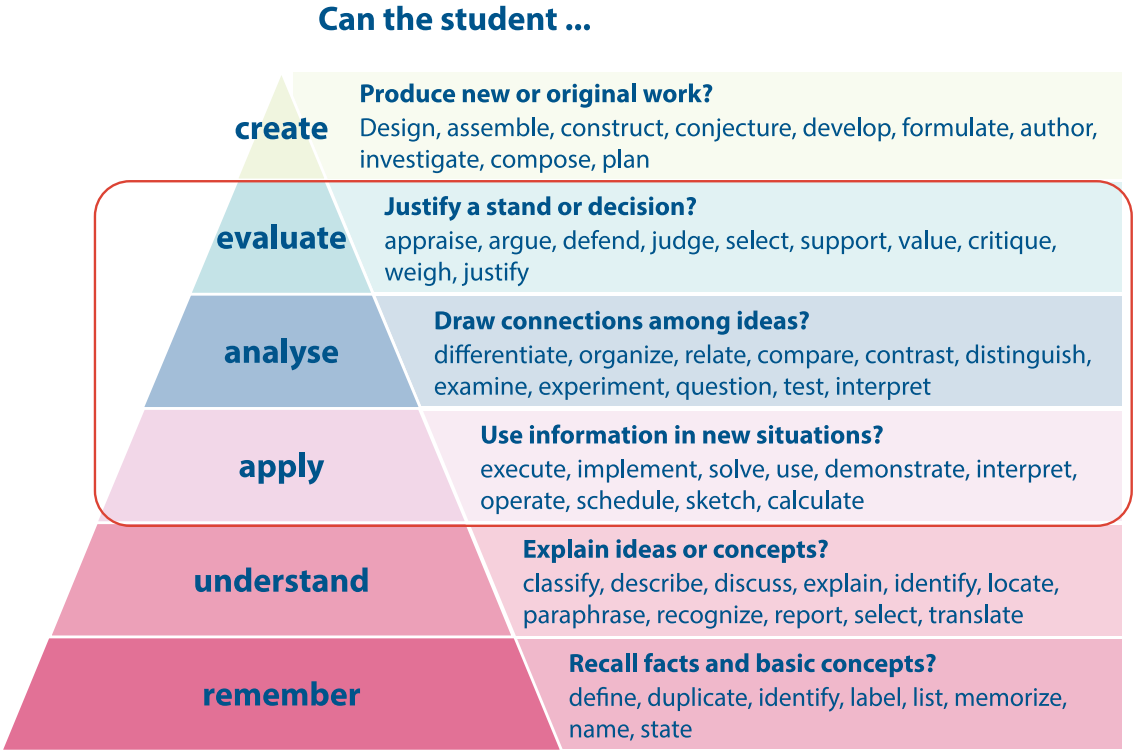
IB assessments are designed to measure the higher-order thinking skills of **evaluation** and **analysis**, not simply knowledge recall.

This point is made by Alec Peterson (2003), whose views shaped the educational philosophy of the IB. He stated that "what matters is not the absorption and regurgitation either of facts or of predigested interpretations of facts, but the development of powers of the mind or ways of thinking which can be applied to new situations and new presentations of facts as they arise". Sugata Mitra, addressing the IB Heads Conference in 2011 in Singapore, took Peterson's perspective a step further, arguing that the immediate and comprehensive availability of knowledge through the internet means that knowledge itself has little value. Rather, Mitra argues, it is the ability to analyse, interpret and create knowledge that is required by 21st-century citizens (Mitra, 2011).

IB assessments give significant attention to higher-order thinking skills, for example, by setting performance expectations in assessment tasks based on established taxonomies, such as Bloom's taxonomy (Bloom et al., 1956; Anderson, Krathwohl, 2001). Indeed, Bloom's taxonomy of educational objectives provides a useful framework to express the diversity of skills and cognitive processes required to successfully complete assessment tasks and earn full marks. Bloom's higher-order skills certainly require the use of a different kind of assessment. Students' skills of analysis, synthesis and evaluation can only be properly gauged by requiring them to analyse, synthesize and evaluate at some length. Performance assessment is the only realistic means of measuring student achievement in these areas, and because the outcomes of such activity cannot be tightly prescribed, these assessments must be relatively unstructured and open-ended, allowing for many diverse but correct responses.

Figure 2

Bloom's taxonomy



The ability of IB assessments to recognize and reward a student's performance in these skills is essential if they are going to be meaningful, despite the challenges this presents to reliability and other aspects of validity. Tests that only reward the recall of knowledge, concepts and routine techniques are not fit for purpose within the goals of an IB education as they do not adequately address these critically important higher-order thinking skills.

The learner profile and skill development

> Education today is much more about ways of thinking which involve creative and critical approaches to problem-solving and decision-making. It is also about ways of working, including communication and collaboration, as well as the tools they require, such as the capacity to recognize and exploit the potential of new technologies, or indeed, to avert their risks. And last but not least, education is about the capacity to live in a multi-faceted world as an active and engaged citizen. These citizens influence what they want to learn and how they want to learn it, and it is this that shapes the role of educators.
>
> (Schleicher, 2016)

It is increasingly claimed that the skills required in the 21st century are fundamentally different to those of previous generations—though some would argue that the inquiry approach that underpins these 21st-century skills has been valued since Socrates. Nonetheless, there is general agreement on the importance of providing students with a wide range of attributes to prepare them for life. See, for example, the arguments made in Llewellyn (2014).

While there are many ways to categorize these skills, including the Organisation for Economic Co-operation and Development's (OECD's) 21st-century competencies, RAND Education and the National Research Council (NRC) framework among others, at the IB, these competencies are defined and described within the learner profile.

Not all aspects of the learner profile are appropriate to measure through summative assessment, but several are encapsulated within the concept of higher-order thinking skills. Good assessment recognizes the importance of these characteristics and, through formative and summative opportunities, creates space for the development of these competencies. Examples of this include encouraging ethical approaches to

surveys and experimentation in psychology and the sciences, the extended exploration of an ethical dilemma in the reflective project, supporting appropriate peer review and introducing unexpected contexts to students.

For more details on the IB's wider approach to student competencies, refer to the material available on the IB website or the relevant programme's *From principles into practice* publication.

International-mindedness and intercultural understanding

The mission of the IB is to help students develop as "caring young people who help to create a better and more peaceful world through intercultural understanding and respect", and "who understand that other people, with their differences, can also be right" (IB mission statement, 2024). IB programmes are also studied by students in many countries and regions (territories) and of many nationalities. There is, therefore, both an **international context** and a focus on developing **international-mindedness** in IB learning and teaching, both of which must be reflected in assessment. An important step in delivering this multiplicity of perspectives is having academic experts, including assessment and curriculum developers, from a wide range of cultural backgrounds. It is important to the IB mission not to obscure differences but to engage with them in a way that allows students to explore them productively and positively.

In some subject areas, cultural diversity can be encouraged through a recognition of varying cultural perspectives and an emphasis on culturally varied content in the curriculum. Examples of this approach can be found in several DP subjects, including biology, chemistry, psychology and visual arts. In the first three of these, the optional structures within each subject allow schools to select course content that will, to a certain extent, suit particular cultural traditions of teaching the subject.

In other subject areas, such as the arts, and language and literature, international-mindedness is encouraged through the recommended and prescribed curricular content. It can also be emphasized through a wider range of IA tasks.

There is more to international-mindedness than just knowledge and understanding of other cultures. **Attitude** and **action** are also important attributes. Attitudes are difficult to assess through normal school assessment.

This challenge is addressed through the non-assessed core elements of IB programmes, such as the community project in the MYP and the CAS component in the DP. As the MYP certificate, the IB Diploma and the IB CP Certificate cannot be awarded without students having completed the programme core, these elements have a significant impact on the overall outcome of IB assessment.

Figure 3

Principled action

Allowing students to choose which questions to answer might be seen as the best way of addressing different international requirements in assessment. However, this approach poses assessment problems in terms of maintaining **comparability** across the options. This occurs most often when there are choices of question, or very open-ended assessment tasks, making it even more challenging to define what "equal demand" means when students come from very different educational backgrounds. In general, it is easier to maintain comparability by setting common tasks that allow students to introduce their own experiences into the answers. In such cases, the challenge falls upon the examiner to maintain a common standard.

Information on comparability can be gained through analysis of student performance. This analysis is discussed further in the section "Grade awarding and aggregation".

Assessment carried out in an international context has additional challenges, in terms of equity, above those normally encountered within a national system. Questions that might be perfectly appropriate in one national setting become inappropriate in another. For example, questions referring to sports, travel, entertainment, historical events or even the weather must be prepared and treated very carefully. It might seem that the only way around this problem is to prepare examination questions that are devoid of all sociocultural context. However, to do so would not only make examination questions quite limited; it would also contradict the philosophy of IB assessment, and good assessment practice in general, which aim to ensure validity through context-based tasks. Contextualized work and assessment are vital to good learning.

There are two possible ways to address this dilemma. In the first approach, appropriate contextual information can be provided to students through specification in the subject syllabus content, by providing case studies on which questions are based, or even in the examination content itself.

The second approach is to use more open-ended assessment questions and tasks that allow students to utilize their own context to build their responses. Importantly, when using this approach, the focus of marking must be on deeper levels of understanding, rather than on straightforward knowledge of subject

content, as there may be no common basis of content. This is very much in keeping with the IB's assessment philosophy.

Figure 4

Range of cultural contexts

Even with the application of both of these methods, students may find themselves dealing with assessment tasks featuring sociocultural contexts that are unfamiliar to them. This is also in keeping with the IB's assessment philosophy, which aims to encourage students to be **more open-minded**, **more globally aware**, and more **comfortable** and **competent** at operating in unfamiliar cultural contexts.

A significant proportion of IB students register for examinations in a language that is not their first or best. Nearly all such cases relate to English, because students working in French or Spanish (the other two languages in which IB assessment is most often conducted) tend to be fluent speakers. Accordingly, during assessment authoring and editorial reviews, considerable care is taken to ensure the wording of questions does not disadvantage additional-language speakers. This is an important priority during the examination editing process.

The role of technology in IB assessment

- Students use technology widely, including in their social life, for their learning and to write their essays. They will go on to use technology extensively in their further education and careers. Using technology in examinations is, therefore, authentic to their experience of the world.
- The IB uses technology to deliver higher-quality and meaningful assessment. The use of technology does not drive the choice of approaches to assessment.
- It is important to separate the impact of technology to support expert examiners (eMarking) and using technology to create meaningful assessment for students.
- The technology that students use for assessment must be familiar to them from their classrooms.

Digital assessment

Digital assessment is assessment taken on a computing device. This could include many different styles of question and stimulus, according to what is being assessed—from media-rich, highly interactive assessments to more traditional essay-style examinations.

In practical terms, digital assessments require each student to work on an individual device in an examination. The current model employed in the MYP is that an internet connection is recommended but not required, as the examinations are loaded on to the computing devices beforehand. The examination software "locks down" the device while it is running and records any occasions when the software is interrupted for any reason (with the invigilator being able to restart in cases of genuine disruption). It is strongly recommended that MYP digital examinations are completed online, as this ensures online capture of student responses at regular intervals and eases the administrative burden for schools. While an overwhelming majority of schools and students do utilize the online functionality for MYP, it is not a compulsory condition for sitting the examinations, and an offline method is supported by the IB. An internet connection is required to run the DPCP digital examinations, with students accessing the examinations via a lockdown mode that limits their access solely to the examination.

There are significant advantages to digital assessment for testing what the IB really wants to test, rather than just what is possible within the limitations of a paper-based examination. These advantages are described in detail in the section "Benefits of digital assessment to validity", and include being able to offer video material, allowing students to use common word-processing tools when writing essays, and supporting accessibility, for example, by giving students control over colours and font size.

Digital assessment will continue to evolve. As technology develops, the IB will be able to develop its assessments to match. Currently, the IB is only considering students using keyboards and mice or track pads, but within a few years, touch-screen technology may become the norm and open up more possibilities. Looking much further ahead, what role could virtual reality play in assessments? It is possible to imagine foreign-language assessments being more authentic in such an environment, while science-fiction films provide an image for how technical tasks (science questions) could be managed through virtual-reality laboratories.

With the IB's ongoing commitment towards digital transformation across its programmes, new terminology will inevitably emerge that impacts the wider assessment community. A glossary of IB assessment terms, including terms relating to digital assessment, can be found at the end of this publication.

Teaching using technology

Technology is changing the way students are taught, whether through the use of interactive whiteboards in the classroom or massive open online courses bringing teaching to a wider audience. Most students produce their essays on a computing device rather than writing them by hand.

The move away from handwriting to using devices creates a disconnect between assessment and teaching practices. For many students, the experience of spending two or three hours writing with a pen is almost unique to taking examinations, and this significantly undermines the authenticity of the assessments.

With the increased use of digital assessments, it is important for students to be familiar with the style of the assessments so that using the interface is not a barrier to them showing their understanding of the subject. While this can be achieved through familiarization tools and mock examinations, the ideal solution is for the students to be using the same tools during teaching.

What does good digital assessment look like?

Digital assessment can improve the quality of assessments by leveraging technology as the means to ask questions that would be impossible in a traditional paper-based examination. At the IB, good digital assessment:

- uses technology to set out more valid, authentic and relevant tasks
- removes administrative barriers and improves delivery
- avoids introducing any new obstacles to completing the assessment.

eMarking

The term "eMarking" is used to describe how the IB marks work submitted for assessment using computing devices to display the scripts and record the marks. At the IB, examiners are no longer sent paper copies of examinations or IA work, other than in exceptional circumstances linked to access and inclusion arrangements. Paper-based work is scanned into an electronic format and examiners can access it via the internet through specialist marking software.

eMarking has a number of major advantages for the speed and quality of marking.

- The IB always has the work submitted for assessment; if, for any reason, it needs to be appraised by a second examiner, it is instantly available and does not need to be posted around the world.
- The IB can anonymize the work to reduce the chance of examiner bias in marking.
- The IB can implement a more rigorous quality-control process and can check marking standards during marking.
- Work submitted for assessment by students from one school can be randomly shared between all examiners rather than all being marked by one examiner. This allows the IB to carry out additional quality-checking processes.

eMarking has been used by the IB since 2010, and does not require students to submit digital work. Currently, most examinations are handwritten and then scanned into a digital form.

Risks associated with digital assessment

In developing digital assessment, the IB is aware of several significant challenges that must be overcome in order to ensure fairness for students and schools. The following sections highlight only a few of the key risks the IB is managing. More detailed discussions will be included in individual sections of this publication and in other IB publications focused on digital assessment.

Burden on schools

The IB recognizes that the use of technology is an integrated part of students' worlds today, and this should include their experience of assessment. However, the IB also recognizes that schools need time to change their processes, and that digital examinations require different arrangements from paper-based

examinations. The IB will manage the introduction of DPCP digital examinations to all schools to keep pace with technological advances, while not placing unreasonable expectations on schools to adapt, whatever part of the world they are in.

Supporting a gradual transition to digital assessment

The IB will work with the IB community to offer timescales to allow all those schools who want to take advantage of the opportunities offered by technology for better quality assessment to be able to do so. The IB believes that the potential of technology should be utilized intelligently to improve the quality of learning and teaching, and, therefore, assessment.

Technology failure

There is no acceptable level of technology failure for students taking examinations. Every student must have a smooth and uninterrupted experience. The IB has a comprehensive familiarization, testing and compliance process that is available to schools before the day of examinations. The IB's experience is that most issues occur as a result of school infrastructure; through the familiarization, testing and compliance process, these can be identified and resolved well in advance of examinations.

Security

Digital examinations need to be at least as secure as paper-based examinations. The IB's assessments are designed to lock down all other functionality of the relevant device while they are being taken. Where assessment material is sent to schools ahead of examinations, as in MYP eAssessments, these files are encrypted and password-protected.

While this remains an area the IB will continue to strengthen as more sophisticated approaches are developed by the industry, it is also aware that current paper-based processes are subject to their own security risks.

Technology for the sake of technology

The IB is committed to using technology to improve the quality of what can be assessed, not using it for the sake of using it. The IB will continue to articulate how it has added value to the validity of its assessment through the use of technology, and the assessment development process will reflect this.

Bias

The IB recognizes that certain students will perform better as a result of being able to use computing devices to answer their assessments. However, the IB also recognizes that paper-based examinations are not free of bias: students who find it difficult to write with a pen for a long time, or who are slow at writing, are already disadvantaged.

As a principle, the IB does not intend to measure students' skills in using a pen or typing, and it will design assessment tasks that do not give advantage to those who can type or write quickly.

The IB is paying close attention to the published research on such "device effects" to ensure it can meet this principle.

Changing standards—up or down

IB grades have a meaning, and this meaning will be protected through the expert judgement of the IB's senior examiners and supported by outcomes data. For example, it may be easier for a student to create a well-structured essay when they can cut and paste paragraphs as they go along. The IB will make sure that this is taken into account when setting grade boundaries, so a student equally comfortable with paper-based and digital examinations will obtain the same outcome—one that reflects their understanding and analytic ability in the subject.

As a result of being able to test new traits through digital assessment, including traits that the IB has always valued but has not previously been able to assess, the IB may need to adjust what it expects students to demonstrate in the assessments—see the section "Standards" for more information. However, the IB will

only do this in a way that upholds its long-standing goals and reflects the qualities described in the current grade descriptors.

Further reading

For more information about digital assessment at the IB, please refer to the following resources.

- The "Transition to Digital Examinations" page on the PRC
- *MYP eAssessment playbook*
- *MYP eAssessment: Introduction for school leaders and teachers* (video)
- *MYP eAssessment Ready! Introduction*

Assessment in education

- In education, assessment has often been categorized as either formative (assessment **for** learning) or summative (assessment **of** learning). Today, there is a strong movement for assessment **as** learning.
- The type of assessment used in any given context must depend on how the outcomes will be used or the purpose of the assessment.
- Assessment can influence teaching practices and must be designed so that any backwash effect—or impact that the assessment has on learning and teaching—is positive.

Defining assessment

As set out in *What is an IB education?* (2019), the International Baccalaureate (IB) follows a constructivist approach to learning that offers an active role for the student and also recognizes the importance of context to effective learning (Murphy, 1999). If assessment is to support effective learning and teaching, then it must be designed around this constructivist learning theory. For more research relating to this concept see Shepard (1992), Wood (1998), Black (1999), and Lambert and Lines (2000).

"Assessment" can mean any of the different ways in which student achievement is collected and evaluated. Common types of assessments include tests, examinations, extended practical work, projects, portfolios and oral work. Sometimes, assessments are carried out over a prolonged period; at other times, they take place over a few hours. Assessments, depending on their type and purpose, can be judged by a teacher, a peer student, the student themselves or an examiner.

Figure 5

Examples of different types of assessments

Assessment tasks that test how good—or competent—a student is can use either a **compensation** model or a **mastery** model (see figure 6).

The **compensation** model is used in most external examinations, and it allows a strong performance in one area to mitigate for a weaker performance in another.

In contrast, a **mastery** model of assessment requires a minimum attainment—or mastery—in each part of an assessment.

An examination consists of three questions, each marked out of 10.

The pass mark is 15, and this could be achieved by a student scoring of 5, 5 and 5 on each question, or by a student scoring 10, 3 and 2. In the second case, a perfect score in the first question compensates for weaker scores in the other questions.

Taking the same examination example (where an examination consists of three questions, each marked out of 10), if a pass mark of 5 was required in each question, the first student would pass, while the second student would fail. A student achieving 10, 10 and 4 would also fail.

Figure 6

Comparison of compensation and mastery models

Compensation model

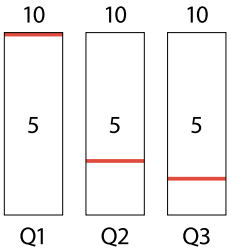
Number of questions: 3

Number of marks (each): 10

Minimum pass mark (total): 15/30

Applying a compensation model:

assessment **passed** based on the total across all questions being equal to or greater than the minimum pass mark.



Mastery model

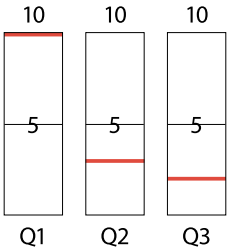
Number of questions: 3

Number of marks (each): 10

Minimum pass mark (per question): 5/10

Applying a mastery model:

assessment **failed** based on an insufficient mark on questions 2 and 3.



Mastery model assessments are often used in more career-related or workplace settings where it is not appropriate to be very good at one element but poor at another. For example, when making an item of clothing, there is no degree of design expertise that can compensate for not having basic sewing skills.

The IB employs a range of assessment tools. These include examinations intended to be taken at the end of a programme, as well as a variety of other assessment tasks (research essays, written assignments, oral interviews, scientific and mathematical investigations, fieldwork projects and artistic performances) spread over different subjects and completed by students at various times under various conditions during their course.

Approaches to assessment

Assessment can be used for a variety of purposes. The intended purpose for a given assessment will have a major impact on how it is designed. Traditionally, there have been two broad types of assessment: formative and summative.

Formative assessment has the most direct link to the way students learn and is sometimes called "assessment **for** learning"; summative assessment is sometimes referred to as "assessment **of** learning". This underestimates the major impact of summative assessment on what is actually learned in the classroom. All

assessment should support appropriate learning. Summative assessment is not just an activity conducted after learning has taken place, but should be designed to have an integrated role in learning and teaching.

The aim of formative assessment is to provide detailed feedback to teachers and their students on the nature of the students' strengths and weaknesses, and to help develop students' capabilities. Types of assessment, such as direct interaction—in the form of a discussion between teacher and student, for example—are particularly helpful here.

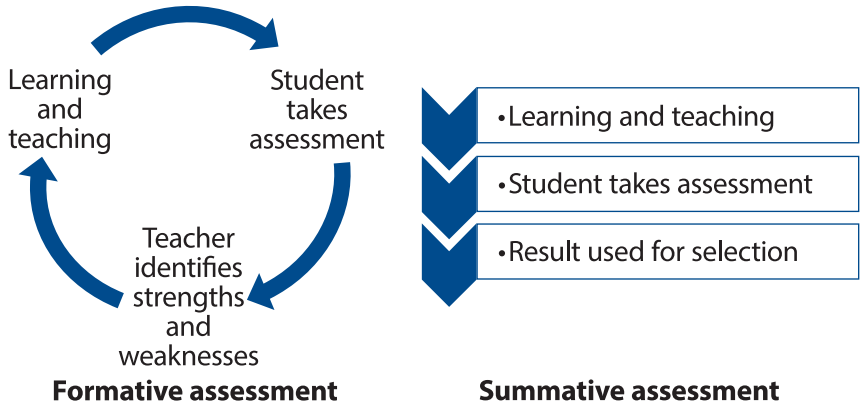
Vygotsky (1962) describes the teacher as a supporter rather than a director of learning, so they should make use of assessment tasks and instruments that help the student work in what he refers to as the "zone of proximal development". This is the range of achievement between what the student can do on their own and what the student can do with the support of the teacher.

The concept of "scaffolding" was formed by Wood et al. (1976). In this framework, the teacher provides the scaffold for the construction of learning, but only the student can do the constructing. The intention of the teacher must be to set formative assessments that are at just the right level of challenge for the student, and to keep adjusting that level as the student progresses.

In contrast, summative assessment focuses on measuring **what** the student can do. It is typically used to demonstrate the completion of a training programme, or readiness to progress to the next stage of education. While formative assessment is interested in **why** a student does something, summative assessment wants to know whether they did the **correct** thing. While this may seem less useful than the question of why, the purpose of summative assessment is to make a judgement about the student, not to inform future teaching. The section "Describing student success in summative assessment" explores the need for summative assessment further.

Figure 7

Possible uses of formative and summative assessment



In formative assessment, it is more important to correctly identify the knowledge, skills and understanding that students have not yet developed, rather than to accurately measure the level of each student's achievement. The balance between these priorities is called "validity" and is discussed in more detail in later sections. This balance between the student's attainment and the quality of feedback is reversed in summative assessment, where the outcomes of the assessment will be used to make decisions about the student, often around competitive selection for employment or educational opportunities, but also to support further teaching.

Any analysis of different national assessment systems will quickly reveal a wide variety of assessment techniques and approaches. All of these systems have their strengths and weaknesses in relation to technical, resource and time considerations, and in their impact on the country's education system. Even if it were possible, in a given context, to start completely afresh in devising an assessment system, there is no universal best technical practice that could be adopted. Instead, the choices made in devising assessment systems inevitably reflect the values and priorities of the broader social context in which they are made. For more research in this area, see Cresswell (1996) and Broadfoot (1996).

It is also important to recognize that summative assessment is increasingly being used as a measure of the quality of teaching, which adds a further dimension about why assessment is undertaken: for the benefit of the education system, rather than the student.

Backwash effect and learning

> What is needed is a process of assessment which is as valid as possible … Yet such a process must not, by its backwash effect, distort good teaching, nor be too slow, nor absorb too much of our scarce educational resources.
>
> (Peterson, 1971)

In this quotation, Alec Peterson, the founder of the IB, recognizes the risk that the way in which assessment is carried out can influence the approach taken to learning. As Surgenor (2010) writes, this can have both negative and positive implications on the learning and teaching experience for students.

In a 1971 paper, Snyder proposed that students create their own understanding of the curriculum based on implicit and explicit messages about what counts in assessment, which is described as the "hidden curriculum". This can lead to students understanding how to pass a subject, but not understanding the subject itself. Another way of understanding this is presented by Gibbs in the following quotation.

> When I retook the exam I just concentrated on passing the exam. I got 96% and the guy couldn't understand why I failed the first time. I told him this time I just concentrated on passing the exam rather than understanding the subject. I still don't understand the subject so it defeated the objective in a way.
>
> (Gibbs, 1992, p. 101)

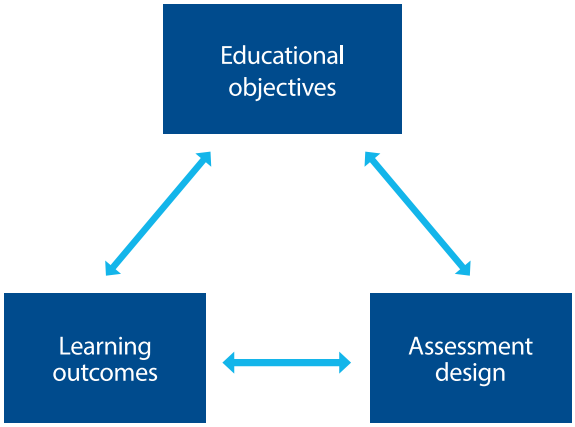
This concept of backwash is also expressed in the adage that "if it is not tested, it will not be taught". This highlights that assessment and teaching cannot be considered as independent of each other.

There are various reasons an assessment can fail to support the educational objectives it is trying to encourage students to evidence. The most likely is that the learning outcomes on which an assessment is based do not effectively reflect the original purpose of the education. For example, a vocational course to train elite athletes might have learning outcomes centred on understanding the rules of athletics. Another common problem is that while an assessment might cover every intended outcome, most of its testing might still focus on tasks that are easy to test but not particularly important to the overall educational goals.

This interdependency of assessment and educational purpose is expressed in figure 8, adapted from Furst's paradigm (1958), in Frith and Macintosh (1984). A disconnect between any two of the three elements it illustrates will almost certainly lead to poor-quality assessment.

Figure 8

Relationship between objectives, outcome and design



Validity

- The term "validity" refers to assessment being **fit for purpose**. It is a complex concept with many aspects.
- Assessments are used for many purposes.
- An assessment may be valid for one purpose, but not another. For example, a spelling test will not also measure fluency in a language.
- It can be argued that it is the purpose for which assessment outcomes are used that is either valid or invalid, rather than the assessments themselves.
- The IB's first concern is whether a programme is valid, followed by whether elements of the programme (such as individual courses) are valid, and finally whether an individual assessment is valid.

Defining validity

What is the purpose of taking examinations? What are examinations for? These questions lie at the heart of what it means for an assessment to be "fit for purpose" or to "measure the right thing"; often, the various answers to these questions are in conflict with each other. It is also the case that not all purposes of an education might be tested by any particular examination; indeed, some purposes might not be possible to test.

For example, suppose that the following four possibilities are given as the purpose of an assessment in mathematics.

1. To recognize what the student understands after their course of study
2. To act as a means of selection for further study or work
3. To give an indication of future success
4. To reinforce the teaching of the curricular goals of the programme

Even these simple goals are difficult to reconcile with one another. If future success in mathematics depends more on calculus than on geometry, should the assessment focus on calculus? How does this relate to the first objective of recognizing what the student knows (in geometry)? What happens when two or more of these goals result in a conflict in how an assessment should be designed?

These four possible purposes are in no way definitive or exhaustive. Newton (2007) sets out a number of examples of possible uses for assessment results.

• Social evaluation uses	• Licensing uses
• Formative uses	• School choice uses
• Student monitoring uses	• Institution monitoring uses
• Transfer uses	• Resource allocation uses
• Placement uses	• Organizational intervention uses
• Diagnosis uses	• Programme evaluation uses
• Guidance uses	• System monitoring uses
• Qualification uses	• Comparability uses
• Selection uses	

Despite this array of conflicting uses, there is still a need for individuals and institutions to make decisions about people in a range of contexts; if assessment results are not available, they are likely to use other, less well-designed and unbiased ways of making these decisions. Underscoring this point, Cresswell (1986) stated the problem as follows.

> Clearly, if the other criteria are less reliable than the examinations, greater reliance on them will lead to less reliable selection decisions.
>
> (Cresswell, 1986, p. 42)

This concept of a qualification (or part of the qualification, such as an individual assessment) being "fit for the purpose for which it is intended" is broadly what is meant by validity.

Validity is often used as an essential term in relation to the quality of an assessment, but there is a long academic history of trying to define its meaning accurately (Newton, 2012). As an added complication, there is a narrower concept of validity, often used together with **reliability**, that describes the extent to which an assessment actually measures what it is stated to measure. To avoid confusion, this publication refers to this second concept as "construct relevance". See the section "Construct relevance and authenticity".

Determining a given assessment's validity is ultimately a matter of judgement. This judgement is based on the systematic collection of evidence to support the interpretations of test scores for the proposed uses of tests. The conflicting demands of an assessment identified earlier in this section will be dealt with in more detail later, but there will always be a compromise between these demands. The important question is whether the available evidence suggests that the assessment is sufficiently fit for purpose to be useful. When talking about validity, therefore, it is good practice to talk about the strength of this evidence—or the strength of the validity argument being made—rather than a simple declaration that an assessment is valid or invalid. Whether something is valid therefore is a judgement based on evidence. Validity is contextual and purpose-driven as opposed to a binary attribute. Newton and Shaw explain that that we must consider the fitness for purpose of an assessment before we make a claim about validity, and that "validation is seen as the process of developing and appraising the strength of an argument concerning the interpretation and use of test scores" (Newton and Shaw, 2014, p. 3). Validity therefore is not the property of the assessment itself, but of the inferences that are drawn from assessment results.

Finally, validity is not something that is singularly achieved during the design of an assessment. Rather, it is continually developing during an assessment's life cycle. Similarly, the validity argument is not made at the start of the life cycle, but is continually added to and refined during the assessment design process.

Creating a validity argument

Validity is not a simple objective concept; it is a balance between competing issues. One cannot "prove" validity. Rather, one can only put together a compelling argument as to why the decisions taken about a course or assessment have made it meaningful and fit for its intended purpose. Consequently, it is not the assessment (or result) that is valid, but the purpose to which it is put. Newton (2012) reflects that "the use of a given assessment procedure for a specific purpose (that is, to make a specific decision) is valid if its interpretive argument is sufficiently strong".

For similar reasons, the evidence for a validity argument should arise naturally out of the processes of developing, monitoring and delivering a programme, and should reflect the discussions and decisions that were reached during these processes. An example in assessment would be that the judgements behind what questions to ask in a particular examination demonstrate that the curriculum has been covered appropriately.

The essential element of any validity argument is to have an appropriate structure that covers all the aspects considered to be important. Within the IB, this is represented by a series of questions for which evidence is gathered.

Maintaining validity

Validity is not something that is decided or "proved" when an assessment model is designed. Rather, it develops with the assessment during its life cycle and, indeed, for as long as decisions are made based on its outcome.

For the IB, different evidence to demonstrate that a given assessment is valid (that is, different validity arguments) will be gathered many times throughout the life of a particular course.

When a course is developed or reviewed, discussions about the purpose of the course and how it should be assessed to meet this purpose will form the core of the validity argument. In particular, the balance between construct relevance and the other aspects of validity will be considered. For every session, the writing of a new examination or assessment task will generate more evidence, especially in relation to its reliability, fairness, comparability, manageability and construct relevance. In this way, teacher feedback and student performance outcomes will provide indicators of how successful an assessment has been in achieving its objective, as well as providing evidence on the reliability of marking and comparability of grades. This information will feed back into the development of the next set of assessments.

Finally, the evidence gathered from all the assessments for a particular course, and all the courses of study within a particular programme, provide the basis for decisions for the next review of the course and programme.

> Evaluating test validity is not a static, one-time event, but a continuous process.
>
> (Sireci, 2007, p. 477)

Elements of the validity chain

- There are many different elements of validity. They can be thought of as a chain, whereby if one link is broken, so is the whole assessment.
- The IB generally focuses on five elements of validity: reliability, construct relevance and authenticity, manageability, fairness and bias, and comparability.
- These five elements are often in tension with each other. When considering whether an assessment is valid, one must consider the main purpose of the assessment to determine the relative importance of each element.
- The IB places the highest emphasis on creating programmes, courses and assessments that demonstrate construct relevance.
- For assessments, this means the IB focuses on providing meaningful tasks and questions that test higher-order thinking skills, rather than examinations that are easy to mark reliably.

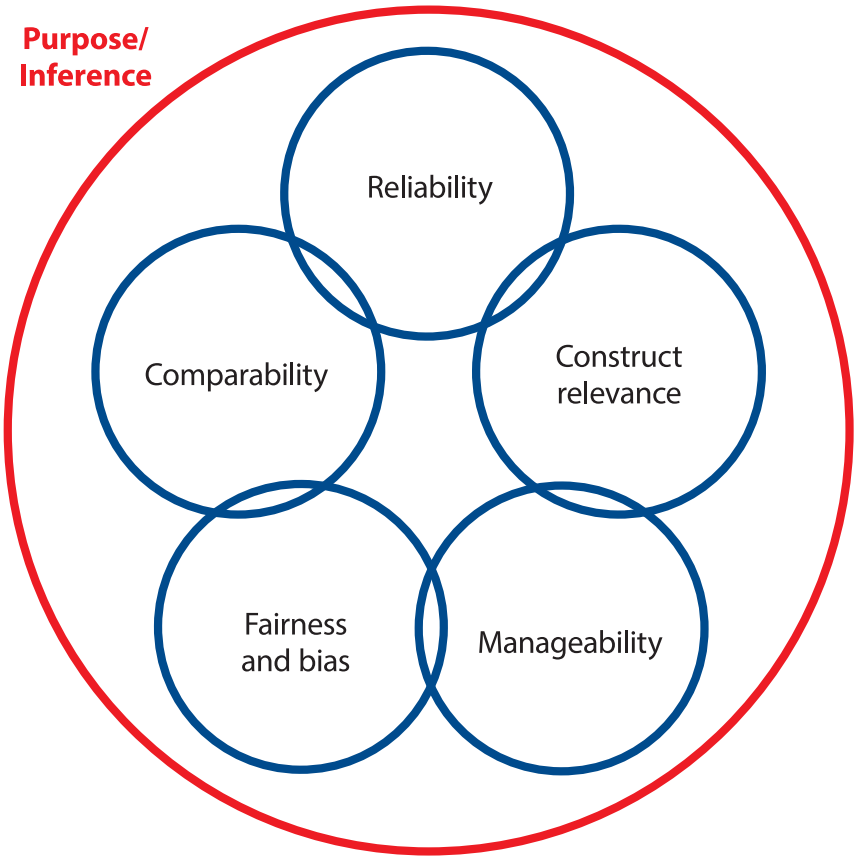
Validity and reliability are widely regarded as essential characteristics of any assessment system, particularly a high-stakes one where the outcome is of great importance to the student or the teacher. These characteristics are in fact multifaceted, with different types of validity and reliability.

The complex nature of validity can be expressed through the idea of the **validity chain** (Crooks et al., 1996). Each of the five elements in the chain is individually important, but are not sufficient on their own to make the assessment valid (that is, fit for purpose). For example, an assessment may be very reliable, but may systematically disadvantage one particular group. Alternatively, its task may focus on exactly what it intended to but, because of its duration and requirements, it may also test the stamina of the student and resourcefulness of the school. The following sections discuss each of these elements in more detail; however, while all of the elements are necessary to achieve validity, there is also tension between each of them. For example, while an examination might aim to cover all aspects of the curriculum, this can easily lead to tests that are too long for students. Similarly, contextualizing a test for each particular culture or country in which it is taken (fairness) might raise questions about whether the same test is really delivered in each instance (comparability).

For an assessment to be valid, all links in the chain must exist.

Figure 9

The validity chain



Balancing aspects of validity

In designing an assessment, it is important to balance these competing aspects of validity in the context of the assessment's purpose. It is not possible for a single assessment to achieve the highest standards in each of these elements, so a compromise must be reached. Likewise, some of these elements are fixed during the assessment design, while others—particularly reliability and fairness—evolve during the delivery of an assessment and its marking and grading.

Figure 10

Balancing competing priorities between aspects of validity: fairness and manageability

		
Balance is wrong Unreasonable burden on candidates (manageability)	**Balance is right** Candidates are assessed in a reasonable way	**Balance is wrong** Too much luck is involved in what topic is tested (fairness)
Assessment contains 50 extended tasks (each 20 minutes long) covering all aspects of the course. This means over 16 hours of assessment.	Assessment contains 20 short questions (2 minutes each), 5 longer tasks (10 minutes each) and 2 in-depth tasks (30 minutes each). Each question is on a different aspect of the course. Total assessment lasts 2.5 hours and covers just over half of the possible topics in the course.	Assessment contains one 30 minute in-depth task, testing both knowledge and understanding on one aspect of the course (out of 50). Candidate's final grade is based entirely on one task.

Despite their separate definitions, there is a considerable overlap in how these five elements occur and are managed. Taken together, they form the wider concept of validity in IB assessments.

Finally, the IB places the highest priority on **construct validity**—that is, that its assessments test the traits and abilities that they are intended to test. Even so, it is important to remember that this prioritization cannot be wholly at the expense of the other aspects of the validity chain.

Reliability

Reliability in assessment can be defined as the extent to which a student would get the same test result if the testing procedure was repeated. This is not necessarily the same as the student obtaining the "right" result.

In their introduction to the concept of reliability, Winkley and Cresswell (2011) list areas where reliability is desirable—such as in marking student performance—alongside examples that highlight a list of potential threats to reliability.

1. **Inter-marker reliability.** One examiner might be more or less lenient on particular questions than the next; similarly, the same examiner might be more or less lenient from one day to the next.
2. **Variability in a student's performance.** A student's performance in an examination might vary from one day to the next, particularly if the conditions of the examination change. These conditions might include whether the examination takes place in the morning or afternoon, who is administering the test, how well the student slept the night before, or whether a caretaker is mowing the lawn outside.
3. **Different examinations.** Different questions will appear from one examination to the next, which might test different facets of a student's understanding (tests usually sample from the curriculum because there is not enough time to test everything, and students may choose to revise one topic but not another).
4. **Comparability of results from one year to the next.** This comparability ensures that student results can be meaningfully compared between years.
5. **Differences between examination specifications.** Ensuring comparability over time can be challenging when there are changes to examination specifications and syllabuses.

6. **Different types of assessment activity.** Many qualifications are made up of different types of assessment activity. These different assessment methods present different types of assessment reliability challenges.
7. **Different types of questions.** Students may perform differently depending on the type of questions they are given.

In practice, the IB usually anticipates a student taking the assessment once, and looks for **consistency** in the results of the process. Consequently, the IB focuses on points 1, 3, 6 and 7 from this list.

As an awarding organization, the IB takes steps to increase the levels of reliability in its assessments. Marking reliability is a central aspect of this, and many of the purposes of standardization, the quality model and moderation are to ensure that examiners mark to a consistent standard.

Reliability in marking

The following quick exercise demonstrates the concept of reliability in marking, and how it can work in practice.

Figure 11 provides five excerpts from authentic responses to an IB language acquisition assessment.

Use the markscheme to judge the quality of the language out of eight, and record the marks you would give to each piece of work.

After you have marked each piece of work, view the feedback about the quality of your marking. The feedback is shown after the student scripts.

Markscheme

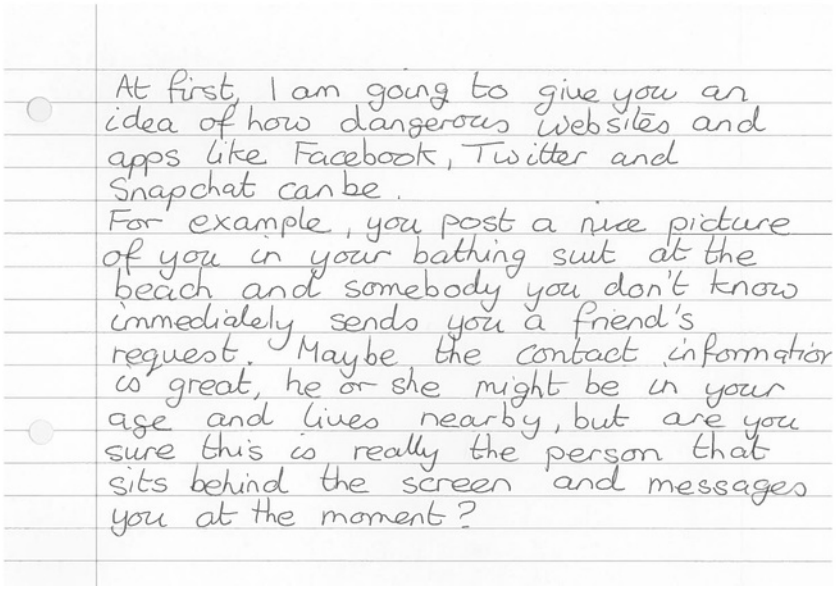
How effectively and accurately does the student apply skills of reflection?

Marks	Level descriptor
0	The student does not achieve a standard described by any of the descriptors below.
1–2	The student states the impact of the project on their learning and whether aims were achieved.
3–4	The student outlines the impact of the project on their learning and whether aims were achieved, and this is partially supported with evidence.
5–6	The student describes the impact of the project on their learning and evaluates the outcome based on the success criteria, which is supported with evidence.
7–8	The student explains the impact of the project on their learning and evaluates the outcome based on the success criteria, which is fully supported with specific evidence and/or detailed examples.

View the scripts in turn, assigning a mark to each one, using the markscheme as your guide.

Figure 11A

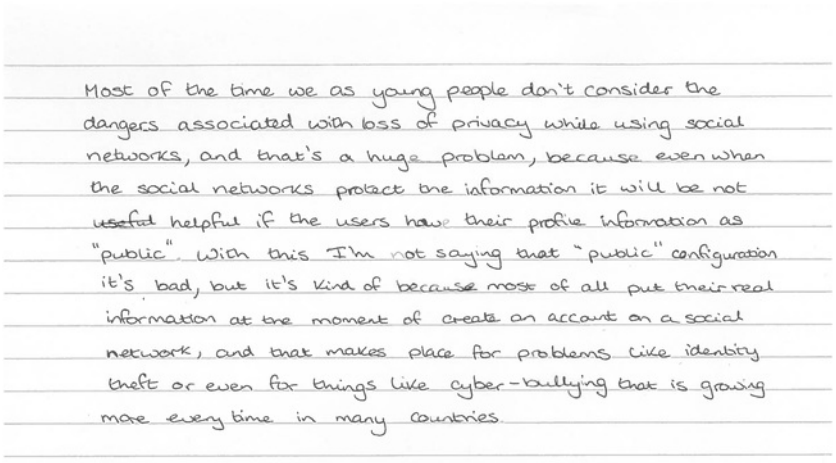
Reliability in marking exercise



At first, I am going to give you an idea of how dangerous websites and apps like Facebook, Twitter and Snapchat can be.
For example, you post a nice picture of you in your bathing suit at the beach and somebody you don't know immediately sends you a friend's request. Maybe the contact information is great, he or she might be in your age and lives nearby, but are you sure this is really the person that sits behind the screen and messages you at the moment?

Figure 11B

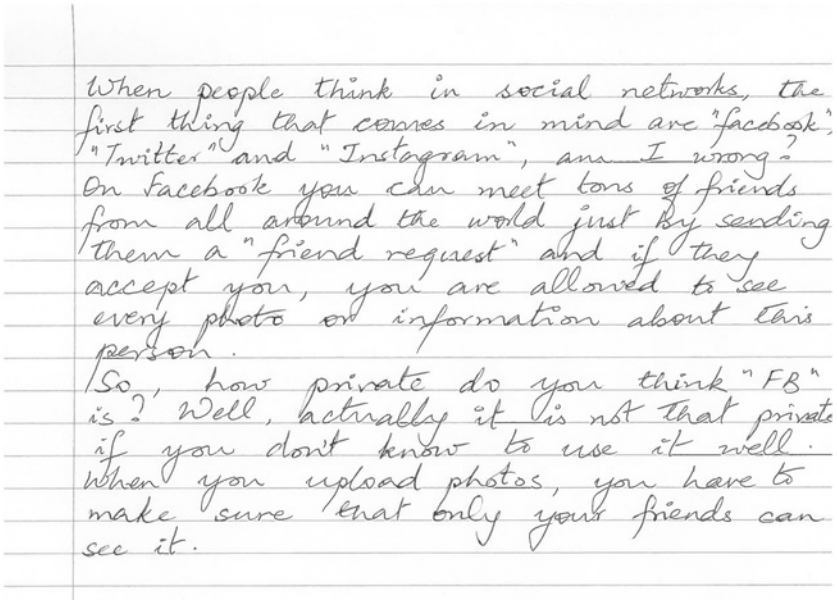
Reliability in marking exercise



Most of the time we as young people don't consider the dangers associated with loss of privacy while using social networks, and that's a huge problem, because even when the social networks protect the information it will be not ~~useful~~ helpful if the users have their profile information as "public". With this I'm not saying that "public" configuration it's bad, but it's kind of because most of all put their real information at the moment of create an account on a social network, and that makes place for problems like identity theft or even for things like cyber-bullying that is growing more every time in many countries.

Figure 11C

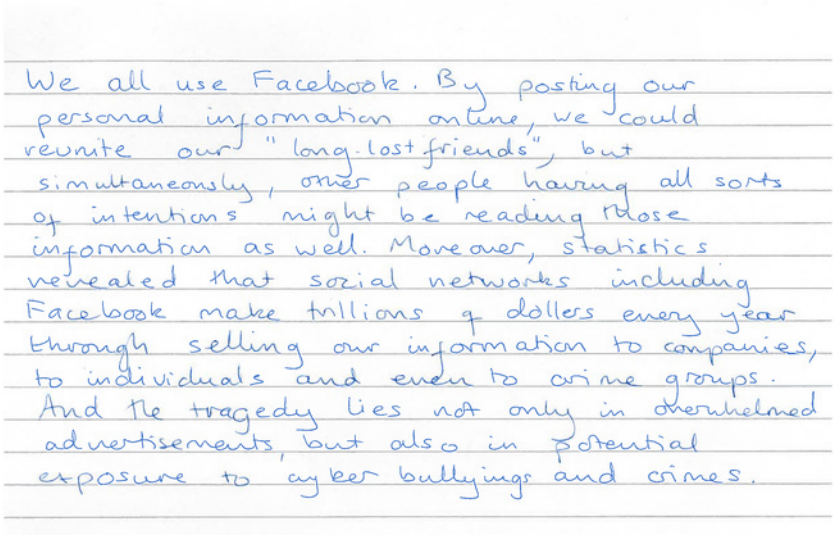
Reliability in marking exercise



When people think in social networks, the first thing that comes in mind are "facebook", "Twitter" and "Instagram", ans I wrong? On facebook you can meet tons of friends from all around the world just by sending them a "friend request" and if they accept you, you are allowed to see every photo or information about this person.
So, how private do you think "FB" is? Well, actually it is not that private if you don't know to use it well. When you upload photos, you have to make sure that only your friends can see it.

Figure 11D

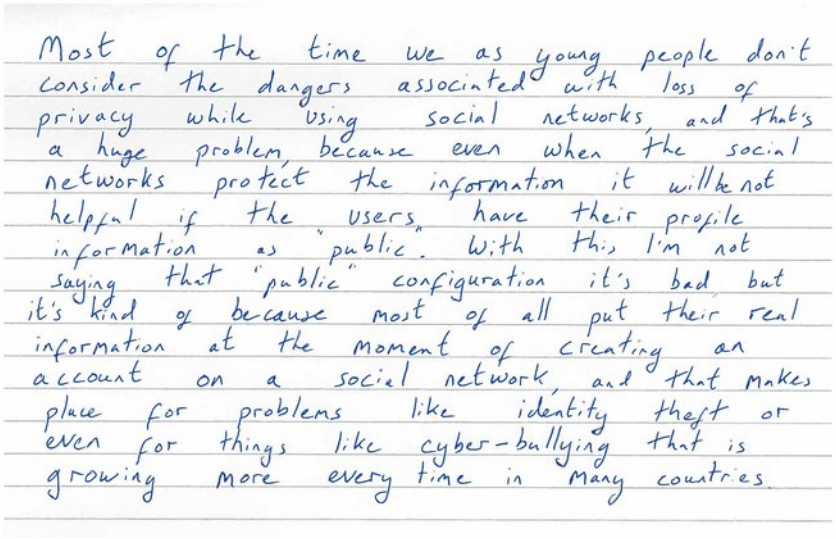
Reliability in marking exercise



We all use Facebook. By posting our personal information online, we could reunite our "long-lost friends", but simultaneously, other people having all sorts of intentions might be reading those information as well. Moreover, statistics revealed that social networks including Facebook make trillions of dollers every year through selling our information to companies, to individuals and even to crime groups. And the tragedy lies not only in overwhelmed advertisements, but also in potential exposure to cyber bullyings and crimes.

Figure 11E

Reliability in marking exercise



When you have completed your marking of each script and given a mark for each student, view the feedback on your marking.

Feedback

Did you notice that student B and student E were the same pieces of work but in different handwriting? Did you give student B and E the same mark?

You may also have found yourself debating which of the two marks in a particular markband to award. How confident are you that you would make the same decision if you were to mark it again?

The point here is not about the quality of the students' work, but how the same person can make slightly different judgements each time they mark a piece of work. Broadly, your view of the quality of the work is the same each time, but the exact mark may vary slightly.

Try this exercise again by marking each piece of work again in a few days' time and see how similar your marks are a second time. (Try not to remember what mark you gave it initially.)

For more information on how the IB ensures reliability in marking, see "Reliability: How we make sure marking is fair" (video).

Consistent outcomes and the "right" outcome

The goal of having a high level of reliability is for students to receive the same (fair) mark whichever examiner looks at their work, not the "right" mark. The question of how good a student's work is relies on professional judgement, and two teachers will often disagree on what mark to award. The point of reliability is that they both provide the same judgement (that of the senior examiner).

This presents a particular challenge when dealing with enquiries upon results. In these cases, the examiner needs to make sure that they provide the same mark that they would have given had this been the first script they were marking, and not be swayed (positively or negatively) by any extra information they now have, such as grade boundaries or the impact on the individual student.

The poor understanding of assessment reliability outside of the education sector has been well documented, but with increasing public discussion of examination results this is a topic that needs greater emphasis.

> As the literature suggested would be the case, the participants found assessment reliability, and in particular measurement inaccuracy, difficult concepts to comprehend.
>
> (Chamberlain, 2010, p. 3)

> Sometimes I think the exams aren't really a fair gauge to a person's ability, are they? They're more about how that person was at the time and how studious they were and all the rest of it.

> Somebody who's not very good at exams could excel later on and become top of their field. (Health sector employee, male.)
>
> (Chamberlain, 2010, p. 27)
>
> While about 63 per cent of teachers and students selected "Any level of error has to be unacceptable—even just one student getting the wrong grade is entirely unacceptable" on one hand, over 50 per cent of them also selected "There's a difference between an avoidable mistake —like a typo on an examination—and something inevitable like inconsistency between two markers", suggesting tolerance for error. This inconsistency may reflect the weak relationship between knowledge about reliability and attitudes to unreliability.
>
> (He et al., 2010, p. 27)

Construct relevance and authenticity

How accurately are we measuring the thing we are trying to measure? The accurate measurement of the skills and knowledge intended for measurement is sometimes referred to as "construct validity". However, to avoid confusion with overall validity, this publication uses the term **construct relevance**.

The idea of an authentic assessment is closely related to construct relevance. **Authenticity** means that testing is done in a way that, as far as is practical, matches the situations in which students would expect to encounter problems in the real world. Assessments that remove tasks from their context, oversimplify or are clearly contrived are examples of inauthentic assessments.

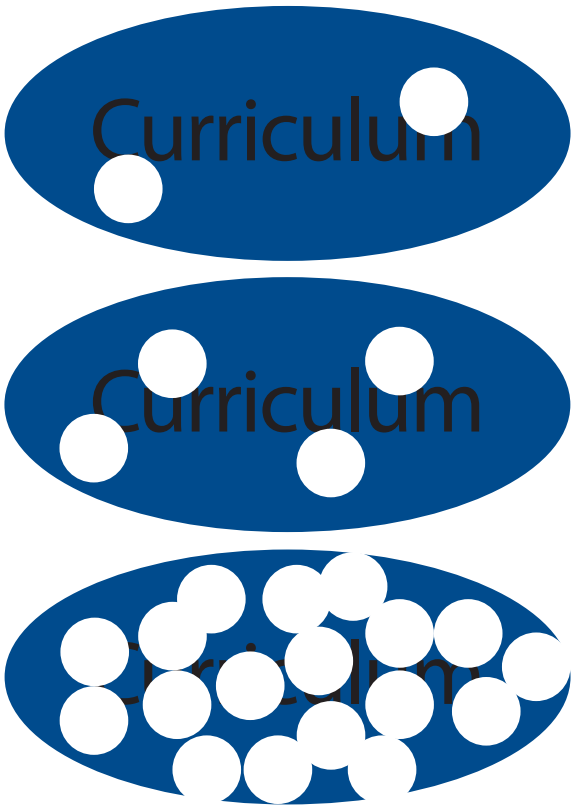
It is easy to create examples of poor assessment design that has a low degree of construct relevance—for example, the ability to write a letter being tested in an oral examination. Often, however, even a widely accepted approach to setting tasks does not truly test what is intended. For example, consider how some traditional literature examinations are constructed. Students may have been taught about the literary devices used in a set text, and the assessment asks them to write an essay related to one of these devices. If the students simply recall everything their teacher said, are they actually demonstrating an understanding of literature?

A clear understanding of what a particular assessment task seeks to allow students to demonstrate, and an inquiring and challenging review of the degree to which it does so, represent best practice in ensuring appropriate degrees of construct relevance. Considering which other skills the student would require to undertake a given task is especially useful. Open tasks, such as essays or projects, are particularly vulnerable to requiring the student to have a high level of competence in writing in order to demonstrate the research or analysis skills that the task is intended to address. This is encapsulated in the concept of universal design for assessment, which seeks to reduce barriers to any students where challenges are not related to what the assessment is designed to evaluate (Dolan et al., 2013).

Construct relevance has a particularly strong relationship with the **curriculum**. In many cases, an assessment will only ask questions based on a small selection of the curriculum material. The question of whether this curriculum coverage is sufficient is one of construct relevance and a matter of judgement. Similarly, the choice of the types of questions to ask is an important part of assessment design, as different question or task types can test different kinds of construct.

Figure 12

Examinations and curriculum coverage



Manageability

The element of manageability has not been comprehensively researched, and there is no single definition or approach to its measurement. In general terms, "manageability" refers to the effort required to take the assessment—for the student, the school and the IB.

Student manageability often relates to the effort required to complete an assessment. For instance, an 8-hour examination is considered to be an unreasonable demand on an 18-year-old. Similarly, the time at which an assessment takes place, and how long the examination period is on a given day for a given student, should also be considered when evaluating student manageability. The IB needs to consider the cumulative burden caused by other IB assessments that students may also be taking as part of their education—this comes back to the principle that the IB considers the validity of its **programmes**, not just the individual **subjects**.

A **school** may be required to provide material for an assessment, which affects the manageability of that assessment. For example, a vocational engineering course might require each student to have an engine to assemble. Depending on the number of students, the number of engines needed could become unmanageable.

Another aspect of school-based manageability is how the students' work is passed to the IB for marking. The need to record or film a presentation represents more of a demand on the school than the submission of a written piece of work.

Finally, the **IB** also considers manageability in terms of the quantity of work submitted for assessment that is reviewed. Attending a three-hour drama production may provide the best evidence of a student's ability, but it is not practicable if the task is to be externally assessed. Offering a wide range of optional questions may also affect manageability, as this requires careful and time-consuming work to establish a common standard between them.

There is often a tension between **manageability**, **reliability** and **construct relevance**. For instance, increasing the amount of assessment undertaken—or increasing the length of a given examination—would provide more evidence of students' understanding of the whole curriculum. This would increase the likelihood that generous and harsh marking decisions will cancel themselves out, thus improving reliability. However, this would also lead to the assessment testing students' ability to maintain their performance in a long examination, rather than the objectives of the assessment, thereby diminishing its construct relevance and manageability for students.

The IB places tight controls on the manageability of assessment, in particular around the total amount of assessment for each course.

Fairness and bias

A test is biased if it gives an unintended advantage or disadvantage to a student. Biased examinations might include, for example:

- a history examination which is written entirely in Latin
- a mathematics examination in which all the questions are based around scoring runs in a cricket match
- an art examination where a painting easel is set up two metres from the floor.

In each case, these tasks may put some students (for example, shorter students in the last of these examples) at a considerable **disadvantage**. In practice, most examples of bias are more subtle than these, but unless care is taken, bias can make a considerable difference to students' results. Putting questions in context without producing bias is a particular challenge; situations that are familiar for some students will be very unfamiliar for others, especially given the international nature of the IB.

"Bias" can be defined as a difference in the outcome of an assessment process that is not related to a genuine difference in the aptitude or achievement being measured. Bias can arise from the way in which the assessments are **delivered**, from the **marking** of an assessment (which becomes an issue of marking reliability) or from the assessment **tasks** themselves.

Bias arising from assessment delivery

The examples from the beginning of this section include bias in delivery—in this case, the unlikely idea of an easel being set too high up for some students. There are many ways in which assessment delivery, and particularly examination delivery, cause bias. The most common relates to the timing of an examination. Examinations held during periods of extreme temperature or pollution in some parts of the world, or during periods when some students are not able to concentrate or prepare properly, would produce bias. This is a particular challenge for the IB, as the conflicting requirements from countries and regions (territories) around the world mean that every possible date is inappropriate for at least some schools.

Another common bias may arise from how an examination hall is set up. Consider two students: one sitting in direct sunlight and another in dark shadow. Inconsistent practices about examination rules might be another example: one student may be kept strictly to time while another is given some flexibility. The IB manages this through clear and consistent rules set out in the *Assessment procedures* for each programme, and by treating a breach of these as maladministration.

Bias arising from marking

Most marking bias is unconscious, rather than deliberate. The human mind is designed to use shortcuts to help decision-making, and these often result in unconscious bias. While the IB does have a duty to mitigate for this bias, there is no blame associated with it.

Bias arising in marking can occur for a number of reasons, such as personal attitude to the neatness, or otherwise, of student handwriting (for example, Hughes et al., 1983), preferential treatment for student gender (where this is known or suspected by the examiner) and undue attention given to factors such as formatting, punctuation and spelling, which may not be significantly relevant in some assessment contexts. This can be dealt with through the training of examiners and checking their work.

Unconscious bias based on factors such as gender, nationality or school is also well documented. To minimize this, the IB seeks to anonymize all work submitted for assessment before it is marked.

Another well-researched area of bias is known as the "halo effect". In these circumstances, the examiner develops a positive opinion about a student if their early answers are of a high quality, and this results in giving them a disproportionate benefit of the doubt with later questions.

Where examiners acknowledge that they cannot fairly and without bias mark a response on a certain topic, the IB has processes available so that the work can be assessed by someone who can mark the work without bias, based on the academic merits of the work presented.

Bias related to assessment tasks

Bias occurs when the difference in outcomes is not linked to what is being measured. For example, using a specific cultural example in which a mathematics problem is rooted, where a lack of knowledge about a certain cultural tradition (not a requirement of the assessment) might hinder understanding and performance (a requirement of the assessment).

In the construction of psychometrically informed tests, any item that is shown to have unusual response characteristics during pre-testing, or which shows substantially different response characteristics for different subgroups of the student population (known as "differential item functioning"), may be regarded as biased and removed from the test. The student subgroups may be defined by gender, ethnicity, social class or language competence, or by any defining characteristic that could be argued to be irrelevant to the construct being tested.

When determining whether a particular assessment or element of an assessment is biased, care must be taken to consider how the task can be explicitly linked to the underlying construct and what the possible factors for introducing bias might be. Basing the decision on purely statistical grounds risks falling into the trap of confusing a biased assessment with one whose purpose discriminates between these groups. Goldstein (1996) and Humphreys (1986) have suggested that it is useful to distinguish between "difference", which is an objectively determined fact, and "bias", which is a judgement about the relevance of the difference.

Black (1999) proposes six of the most common possibilities by which questions might be unfair in their impact on different students.

1. The context in which the question is set may favour students familiar with that context. For example, cultural references relating to the United States (USA) favour those located in the USA compared to students elsewhere.
2. Essay questions on human relationships may favour students from cultural or family contexts in which emotional expressiveness is encouraged.
3. Multiple-choice questions may favour boys.
4. Coursework or project work may favour girls.
5. A question using language or conventions associated with a particular socioeconomic background may disadvantage students from a different background.
6. Some questions may be intelligible only within certain cultural contexts. For example, a question about older people living on their own might be quite unfamiliar in some cultural contexts but familiar in others.

Furthermore, while evidence of gendered bias exists, particularly in the USA, it is not clear which aspects of question format contribute to these findings.

In designing assessments, it is not always clear how one should respond to differences among the students taking a particular assessment. Assessments should be designed so that, by means of a variety of tasks and question types, the overall impact of bias is reduced. Any form of cultural or gendered stereotyping—whether explicit or not—should be avoided. The content of individual questions must be scrutinized to avoid the more obvious categories that are known to introduce unfairness, and pre-testing of questions on samples of different groups among the student population might reveal hidden cases. However, if all biased question types and possible scenarios are excluded, there is little choice left available to assessment

designers and question constructors. The resulting constraints would have a negative impact on the validity of the assessment. Apart from avoiding obvious and unnecessary pitfalls, a balanced approach to assessment design, using different types of assessment task and format, seems to offer the most reasonable solution.

There is also the question of how many differently defined groups of a population require consideration. For example, should account be taken of those students who have different kinds of learning styles? As Hieronymus and Hoover (1986) have stated, if differences in interest and motivation are considered to be biasing factors, all tasks or assessment methods may be said to have a certain amount of bias. For example, passages of text used in language examinations are bound to be of more interest to some students than others. Equity in assessment, which includes the avoidance of bias, is a major issue, particularly in certain countries and regions (territories) where any demonstrable bias in an assessment instrument may lead to litigation. However, the proof of bias, as opposed to difference in performance, is often a matter of fine judgement, linked strongly to the particular social context in which the assessment is conducted.

Gipps and Murphy (1994) conclude *A Fair Test? Assessment, Achievement and Equity* by stating that "there is no such thing as a fair test nor could there be: the situation is too complex and the notion too simplistic". However, this does not mean that assessment designers and question writers should not do everything in their power to reduce the impact of bias and unfairness. Gipps and Murphy also maintain the view that assessment designers should set their goal as equality of opportunity and of access to assessment, rather than the equality of outcome that is engineered by manipulating individual test items according to their response statistics. They question the extent to which it would be justifiable, for example, to bring multiple-choice questions into English examinations to improve the relative performance of boys, since this would distort the validity of the assessment according to widely accepted definitions of the term.

It is widely recognized that lack of fairness in the assessment process is only one factor contributing to inequity in education, and possibly one of the less significant ones. There are many other sources of inequity in education that have a major impact on student achievement, such as:

- differences in the quality of teaching within a school
- differences in the level of resourcing for different schools and in different geographical areas
- differences in the level of social and family support given to individual students.

Any of these could significantly affect an individual student's success in a way for which no assessment process, however fair, could compensate. Smith and Tomlinson (1989), for example, found that school effectiveness was a much greater factor in determining differences in examination results than students' ethnicities, indicating that attempts to adjust assessment instruments to remedy differences in performance by different groups may sometimes be inappropriate.

This kind of consideration formed the rationale behind testing for aptitude rather than achievement, but it has come to be understood that assessment of pure aptitude, ability or potential, separated from social background and educational experience, is not possible. It is also impossible to regard educational achievement in a manner that is independent of social and cultural context. The concept of educational success is defined and measured according to the standards of a particular, delimited section of any given society.

Removing bias and barriers to assessment

A further aspect of bias that must be countered is the potential for an assessment task to discriminate unfairly against any student. This is done by ensuring that the assessment conditions allow for the implementation of necessary arrangements to remove barriers to assessment. This ensures that all students can demonstrate their level of educational achievement on equal terms with other students.

The "Prioritizing fairness for all" section of this publication deals with this idea in greater detail.

Recognizing bias

It is important to keep in mind that bias can be positive as well as negative. If a task is particularly familiar to one group of students, or easier for them to complete, this is still bias. The aim of a fair assessment is to provide an equal opportunity for all students.

This publication has described how bias can be introduced in the design of assessments or in the marking process. While it is very important to think proactively about bias during these parts of the assessment cycle, it is not enough just to think about potential bias. Rather, evidence of bias in student results must be sought, analysing how the student performed on a particular question compared with the examination as a whole.

It is also important to base decisions around bias on evidence, not on stereotypes.

Comparability

Comparability is one of the most complex aspects of validity. Assessment outcomes are frequently used to compare students for selection purposes. Where two students have taken the same examination at the same time, it is possible to be reasonably confident that a student with a grade 7 has performed better on the day than a student who achieved a grade 4, but more complicated comparisons are often made. Such comparisons might be, for example, between:

- two students who achieved a grade 6 in history but answered different questions or took different options
- two students, one who achieved a grade 5 in Spanish literature in May 2025 and one who achieved a grade 5 in the same subject in November 2023
- two students, one with a grade 4 in physics and the other with a grade 4 in chemistry
- two students, one with a grade 4 in mathematics and the other with a grade 4 in geography
- a student in one geographic location with a grade 3 in computer science and a grade 6 in Chinese literature, and a student in another geographic location with a grade 5 in Japanese literature and a grade 4 in biology
- two 15-year-old students, one who achieved an MYP Certificate and the other who took a different awarding organization's qualifications
- two students, one who achieved a grade 6 in Indonesian B at standard level (SL) and the other who achieved a grade 5 in Indonesian B at higher level (HL).

Comparability asks whether two assessment outcomes can be considered equal in some sense. Between subjects this is particularly difficult as the IB is testing different things and then asking if they are of equivalent value.

The concept of validity being for a particular purpose rather than a characteristic of an assessment in the abstract is particularly relevant here. A student's result in music is a better indication of their readiness to become a professional musician than their result in visual arts, but both results might be equivalent in predicting their readiness to study history.

The issue of comparability is made even more complex as each student has their own strengths and weaknesses, and will find it easier to perform in some subjects than in others. This is a particular challenge for the IB as the cohorts of students taking each examination is not the same, particularly where there is a choice of courses available.

The IB seeks to maintain three principles about comparability.

- The standard of work to achieve grades within a subject or discipline is comparable between sessions within a year and between years.
- Grades between subjects or disciplines have a consistent meaning so that different routes to achieve the programme award (for example, IB Diploma, MYP Certificate) are comparable.
- While the IB maintains a commitment to the assessment of higher-order thinking skills, attention is paid to ensuring that grades achieved in other qualifications can still be compared to those achieved in the IB.

Measuring comparability

There are many ways to measure how comparable two assessments are, and many academic papers have been written on the topic. In their review of the literature on inter-subject comparability, Coe et al. (2008)

separate methods for comparing how difficult examinations in different subjects are into two broad categories: statistical methods and judgemental methods.

Statistical methods focus on comparing students' performance in assessments and looking for trends. This is based on the idea that, if two assessments are comparable, then, given a large enough random sample of students taking both assessments, their results should generally be the same on both.

In contrast, judgemental methods use subject experts to look at a number of assessments and give their considered opinion on their relative difficulty. A range of research tools and techniques are used to ensure that they are comparing like with like.

Both approaches are perceived as having serious conceptual shortcomings, and Coe et al. (2008) identify six broad criticisms for each of the techniques.

Table 4

Criticisms of statistical and judgemental methods for comparing assessments

Criticisms of statistical methods	Criticisms of judgemental methods
• Studies may measure factors other than difficulty, such as teaching or motivation. • Multidimensionality—subjects may not have a common trait. • Unrepresentativeness—are the statistics based on an inherently biased group of students? • Subgroup differences—if different groups of students experience different degrees of difficulty, does this challenge the conclusions of relative comparability? • Disagreement between methods—can any one method be "correct"? • Problems of forcing equality—what would be the impact on those students taking the qualifications?	• Breadth of criteria—to be applicable across different subjects, criteria must be very broad and, therefore, imprecise. • Crediting responses to different levels of demand—examiners tend to give more credit to good answers to easy questions than weaker answers to harder questions. • Crediting different types of performance—examiners struggle to compare different types of task, for example, short answers and essays. • Even "judgement" methods are underpinned by statistical comparisons—they are based on the experience of how typical students are likely to perform. • Interpretation and context—judgemental evaluations of student ability or fairness must consider structural differences such as those between single terminal examinations and a series of modular assessments. • Aggregating judgements—most assessments measure several criteria, which must be balanced.

Proponents of either methodology are often highly critical of the alternative approach, and others argue that all the current approaches are fundamentally flawed. To explore this issue, Coe et al. (2008) apply five different ways of measuring inter-subject comparability to England's GCSE and A level qualifications. They conclude that there was a reasonably high level of agreement between the statistical and judgemental measures of inter-subject comparability, and that the differences between them were far smaller than the differences between subjects. They also found that relative subject difficulties were stable between years.

The IB maintains comparability between years and options through triangulation of examiner judgement, statistical analysis and teacher feedback (see the section on "Grade awarding and aggregation"). It also reviews comparability at subject or discipline level through both statistical and expert judgement approaches.

The IB's approach to validity

- The IB believes that construct-relevant and authentic assessment is more important than maximizing reliability.
- The IB believes in a rounded, holistic education. Its priority is for strong arguments of validity at programme level. This is valued more highly than the validity of individual courses or optional routes within courses.

Validity is a complex and multifaceted balancing act between a number of important and conflicting demands. There is no single right way to balance the competing elements of validity; where the balance is placed is ultimately a judgement based on the values of the organization that is developing the assessments.

The IB places a high value on testing **what is important** in a way that **reflects the real world**. The first of these points is captured by the term "construct relevance". That is, IB assessments ask what is really important for the subject, not just what is easy to mark. The second of these points refers to authenticity, which means that IB assessments include meaningful tasks that reflect the way that students might encounter the relevant activities in the real world, rather than being artificial and contrived.

These objectives come at a cost to other aspects of validity, most significantly the **reliability** of the assessment. Meaningful, authentic tasks generally require a large degree of subjectivity in marking, which means accepting larger variations between examiners than in the case, for example, of multiple-choice assessments that have one clear answer and can be marked objectively. Moreover, these tasks also have an impact on the **manageability** of the assessment. The kinds of assessments favoured by the IB are more challenging and time-consuming to create, to administer and to mark. They also require more commitment from students. This increases student workload: for example, in undertaking meaningful research tasks rather than focusing just on performance in examinations.

While the IB accepts that other organizations may choose different priorities in balancing assessment validity, it is confident that its position is appropriate and defensible for externally verified IB assessment.

The IB aims to do more than other curriculums by developing inquiring, knowledgeable and caring young people who are motivated to succeed. As outlined in its mission statement, the IB, through its educational programmes, hopes its students will help to build a better world through intercultural understanding and respect. Moreover, each of the IB's programmes is committed to the development of students according to the IB learner profile, which identifies 10 attributes that the IB values and seeks to develop in students.

The importance of this for assessment is that the purposes of the IB are defined at the programme level, not at the level of individual subjects or disciplines. Therefore, the question of assessment validity must be asked at programme level and must include the rules that govern the award of the overall programme certificate, not just each individual grade.

This does not mean that it is not important to consider each course or even each individual assessment task. Indeed, some aspects of validity only make sense at this level of detail. However, in making any overall validity argument, the IB considers the complete programme of study the student has undertaken.

Most of the principles of assessment that are outlined in this resource apply to the whole of an IB education, but "Section C: IB programme-specific processes" looks at each programme in turn and discusses their unique features.

Benefits of digital assessment to validity

- Today's world is not paper-based; computing devices are part of every aspect of life.
- Paper-based examinations cannot provide moving images or allow meaningful interactions with students.
- The versatility of devices allows students to make the visual and audio modifications they need to access questions, rather than needing to request modified papers months in advance.

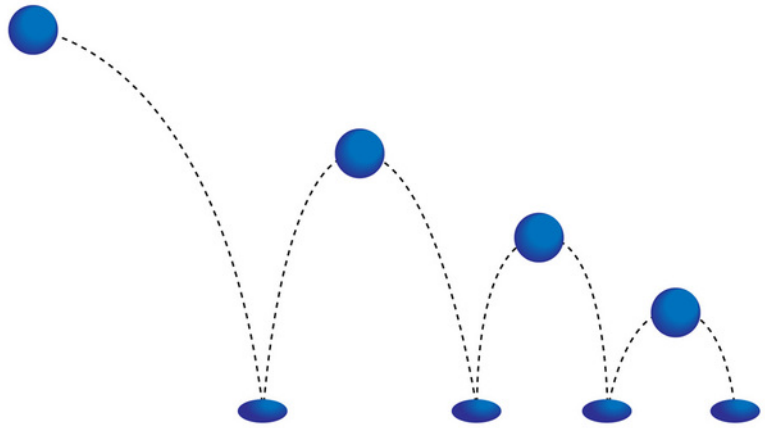
- Despite concerns, digital assessment offers more protection against student academic misconduct and school maladministration.

Email, texts and social media have become far more common than letters as a means of communication, and using a computing device is a routine part of many jobs. If the IB aims for authentic assessments, then it would make sense to incorporate the use of technology and devices in those assessments.

Historically, assessment has been limited in what it has been able to assess through traditional means of testing (for example, using paper and pencil). Within paper-based examinations, students have been limited to responding to a simple stimulus or question without any opportunity to manipulate or interact with the assessment. At the most basic level, digital assessments offer the opportunity to include audio and video stimuli, which can be accessed by an individual student rather than relying on an invigilator playing material for the whole class. In the Middle Years Programme (MYP), digital assessment also allows for animated diagrams and simulations to be included, where appropriate, rather than a picture and a long description of how an object is moving.

Figure 13

Improving the clarity of an eAssessment question through the use of a moving image



As assessment tasks become more sophisticated, technology makes it possible to assess how a student engages with a problem, responds to new information or interacts with a simulation. With intelligent tutoring systems, it may be possible for a digital assessment to respond to prompts from the student, recreating the opportunities offered in teacher-run assessments, such as an oral examination, but without the problems of different students having different teachers. As a result, a key benefit of digital assessment is that it allows for the design of new tasks to assess what the IB actually wants to test, rather than being limited to what can be assessed in a paper-based examination.

Another key aspect of validity is minimizing bias, particularly when inclusive access arrangements are required. The IB regularly processes requests to produce examinations with different fonts or different coloured paper. There are also requests for examination papers produced in Braille.

While digital assessments cannot fully remove these barriers, technology can empower students to select the font size and colour that best suit their need or preference. There are also major accessibility benefits for students using software that enables the use of screen readers. While some students may not be able to use their device to address their access requirements, digital assessment will allow access to a wider range of students without the need for additional support, which can still be used where appropriate.

Among the most prominent concerns about digital assessment—which is also a major threat to validity—is test security. Some critics of digital assessment claim that it is less secure than paper-based examinations; in reality, digital assessments are protected from some of the risks paper-based examinations face, but are also susceptible to others.

Among the more significant benefits of digital examinations is that they are delivered securely, and are only accessible within a window of time defined by the IB, with users unable to access examination content before access is granted by the school. This compares favourably with the risks of sending paper-based examinations, which can be read by anyone, and require schools to keep them secure for several days before the examination.

The IB recognizes that digital assessment presents additional challenges during the examination itself, as students have access to devices that connect to the internet. To reduce the risk of students accessing unauthorized material via their device, IB digital examinations operate in lockdown mode, which prevents access to any other content while they are running, in order to support examination security and academic integrity. Moreover, the ability to record how students answer questions (for example, log files), as well as their final responses, provides opportunities to identify behaviour associated with academic misconduct.

Standards

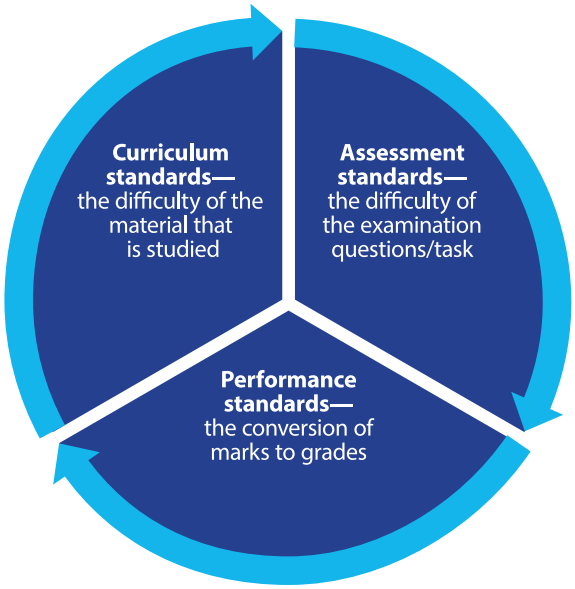
- There are three parts to any definition of standards: **curriculum**, **assessment** and **performance**.
- In formative assessment, the focus is usually on curriculum and assessment standards.
- The IB uses **attainment-referencing** (formerly known as "weak criterion-referencing"), which means using a balance of criteria and the comparison of student outcomes in previous years to set the standard.
- Maintaining standards is as important as setting them.

Three aspects of standards

In the context of assessment, the word "standards" generally refers to how difficult the tasks set for students are. Standards are a core part of comparability. The concept of a standard can be considered in three different ways.

Figure 14

The three aspects of standards



Each of these aspects can be varied to change the overall difficulty of the subject; however, changes to each have different impacts and timescales.

For the IB, **curriculum** changes generally take place during the curriculum review, although it is possible to make them outside the formal review cycle. Fairness requires that the IB gives teachers time to adjust their teaching to adapt to any changes. For the Diploma Programme (DP) and Career-related Programme (CP), this would ideally be two years (the length of the course). For the Middle Years Programme (MYP) it could be longer, given that the programme is five years in length, although typically the assessment only evaluates the last two years.

Assessment standards are dependent on the examination development cycle, typically 12–18 months. Unlike curriculum standards, assessment standards will vary slightly from year to year as it is impossible to design two sets of questions that are of identical demand. Even repeating the same questions as the last

examination will provide a lower level of challenge to students as, to some degree, they will be familiar to them the second time. On occasion, a decision may be made to change the assessment standards of an examination if previous examinations were not performing as they needed to. Depending on the scale of the change required, it may also be possible for an examination that is already in development to be amended. Again, fairness requires that the IB gives notice of the change to teachers, but depending on the scale of the changes this would be considerably shorter than for changes to the curriculum.

Performance standards (that is, grade boundaries) are adjusted every year to balance any change in assessment standards. For example, if an examination is more difficult so that students generally achieved lower marks, the grade boundaries would be moved down so that the same quality of student achieved the same grade as they would have in the previous year. Unlike the other two types of standards, it is primarily examiners, rather than teachers, who need to understand the performance standard so that they can maintain it across sessions. It is possible to change the performance standard at the point at which grade boundaries are set. If the IB were intending to do this, it would usually warn schools in advance so they could manage and adjust student expectations. In such a case, the critical step is that examiners understand what this new standard looks like in order to be able to apply it in future sessions.

In setting the overall standard for an assessment, it is important to balance these three different definitions. While it is possible to set extremely challenging questions on simple material, or to set extremely high performance standards on a simple set of questions, this often results in very poor levels of construct relevance. For example, requiring full marks on a test to achieve a grade 7 requires students to have very high levels of accuracy. A student who has an excellent grasp of the subject but is slightly careless or has poor writing skills is unlikely to obtain a grade 7—is this the intended purpose of the assessment?

The definition of standards applies to both **formative** and **summative** assessment. However, in formative assessment, the setting of the performance standards is usually part of the judgement of the teacher as they decide on the feedback to provide. There is usually far more consideration of curriculum and assessment standards in formative assessment, in order to make sure that a given test is appropriate for the student and will provide useful information to inform future teaching.

Norm-referencing and criterion-referencing

- "Norm-referencing" means rank ordering student performance and awarding a fixed proportion of students each grade.
- "Criterion-referencing" means setting the performance standard according to a description of what to look for in student performance.
- The IB uses attainment-referencing (also known as "weak criterion-referencing"), which means using a balance of criteria and the comparison of student outcomes in previous years to set the standard.

Norm-referencing and criterion-referencing represent two different ways in which a performance standard in assessments can be set and maintained.

Norm-referencing

Norm-referencing is often associated with standardized tests. The principle is to trial a test on a typical sample of students, and then use the outcomes (which, by definition, should be a normal distribution or bell-shaped curve) as a reference scale by which to produce a score for any subsequent student taking the same test. This process of deriving a standard distribution of scores from the initial trial is called "norming".

This technical description of norm-referencing does not necessarily imply that a fixed distribution is applied to every set of test results; the fixed distribution is only used for the original norming. The distribution of scores by subsequent students can vary from this normal distribution.

In practice, norm-referencing often refers to a process where students are put in a rank order according to performance and the proportion receiving each grade is fixed. For example, the top 15% of students would be given the top grade.

Criterion-referencing

Criterion-referenced assessment was first put forward by Glaser (1963). It represented a significant change in setting performance standards, putting an emphasis on measuring student achievement "with respect to a well-defined behavioural domain" (Popham, 1978).

In criterion-referencing, student performance is compared against a predefined description of what is expected at each grade. This is typically done by subject or assessment experts using their professional judgement.

One limitation of this approach is that it is very challenging to create descriptions that are unambiguous and mean the same to all expert judges. Indeed, it has been argued that "no criterion, no matter how precisely phrased, admits of an unambiguous interpretation" (Wiliam, 1993). The outcome of a traditional criterion-referenced test is that mastery of the relevant domain has either been shown or not shown.

In practice, both approaches have severe disadvantages. Strict norm-referencing requires strong evidence that the current test is of the same difficulty as the initial test, while criterion-referencing is subject to the Good and Cresswell effect (1988), in which expert judgement does not accurately take into account the demand of the questions.

Attainment-referencing

The IB uses an approach known as "attainment-referencing" (Newton, 2011), previously called weak criterion-referencing (Baird et al., 2000), which is based upon criteria but recognizes the evidence of the Good and Cresswell effect (1988). In practical terms within the IB, expert examiners are asked to establish a narrow range over which the grade boundaries could lie based on the criteria (grade descriptors), and this is compared with boundaries calculated to match performance from previous years. The grades are set where these two boundaries align; if they differ, there is further discussion to establish how this contradictory evidence can be aligned.

Finally, it is important to keep in mind that criterion-referenced tests and norm-referenced tests differ more in the analysis and interpretation of student responses than they do in the kind of questions set.

Maintaining standards

- Once the appropriate standards have been established, the IB needs to ensure they apply every year.
- Curriculum standards are reconsidered during the curriculum review.
- Assessment standards and performance standards are maintained using a mixture of professional judgement and statistical evidence.

Comparability is an essential aspect of validity. Within the context of maintaining standards, this means that the IB needs to ensure that the same standards apply every year. IB standards are based on the meaning of grades, not marks—the difference between the two is explained in the section "The difference between marks and grades" in this publication.

Curriculum standards are the easiest standards to maintain as they are the same between sessions. Assessments must be true reflections of the whole curriculum, and this is monitored during examination writing. However, external factors can result in shifts in curriculum standards. The classic example is with computing skills. As devices develop and become a more familiar part of everyday life, knowledge and understanding that was once specialized and perceived as demanding becomes commonplace. Therefore, the curriculum standard for the topic has changed despite the content remaining the same.

The IB has a cycle of curriculum reviews to address this issue and to keep content up to date. As a result of these reviews, there is an expectation that the curriculum standards will change—the IB must then balance the assessment and performance standards in order not to disadvantage students.

Assessment standards in IB examinations change every session as no two examinations can be identically demanding. In some education systems, extensive pre-testing is undertaken to establish the demand of each question, and examinations are carefully constructed to ensure a high level of confidence in the degree of difficulty experienced by the students. The IB does not undertake pre-testing of examinations

because of the risk of questions being published before they are taken by students. It therefore relies on the professional judgement of experienced examination setters and scrutineers to create consistent and balanced examinations.

The IB also adjusts **performance standards** based on students' answers to ensure the overall standard is maintained. The two main ways of maintaining performance standards are discussed in the previous section, and the maintenance of performance standards is the purpose of the IB's grade award process.

Describing student success in summative assessment

- While grades represent a very simplified view of student achievement, they allow stakeholders—such as universities, employers, or colleges or schools—to make reasonable judgements around selection.
- If only more complex and holistic information is provided, stakeholders must simplify that evidence to make meaningful selection decisions. They may take less care in doing this than the IB does in setting grade boundaries.
- In IB assessments, the use of professional judgement is important to achieve a meaningful outcome for the student. The IB recognizes that it must support this judgement with objective and unbiased evidence to ensure fairness.

The impact of grades

Consider the knowledge, skills and experience of an excellent chef. The chef has knowledge—of ingredients and flavour combinations that will make up a perfectly balanced dish. They have skills—in selecting and preparing ingredients, in cooking, and in plating and presentation. And they have experience—of using technique to achieve perfection, of judging when something is perfectly cooked, and of presenting a combination of ingredients and dishes to achieve sublime satisfaction at the table.

Now reduce all of that complexity into a single grade out of 7 to decide who is the best chef. The result is almost meaningless. Even increasing the scale to 100—or even 1,000 grades—is unlikely to help; there may be more scope to differentiate, but the fundamental issue of trying to compare different skill sets does not go away.

This is exactly the issue faced within assessment. How can the complexity of a student's knowledge, understanding and ability be best represented in a single outcome? Even if it were possible to precisely capture and summarize all the information about the student, the IB still could not give them a grade that perfectly reflects their talents.

The alternative suggestion, which is often proposed, is not to award a grade at all. Perhaps the IB could provide each student with a personalized description of what they did well and where they were weaker. Such an approach is in keeping with the principles of good learning and teaching, but it has the severe drawback that comparisons between students are very difficult.

The purpose of IB assessment is to support students in being able to signal achievement and progress to further study or work. This means that some kind of selection process is needed by the receiving institutions (often universities for the DP and CP) to determine which students can be given these opportunities.

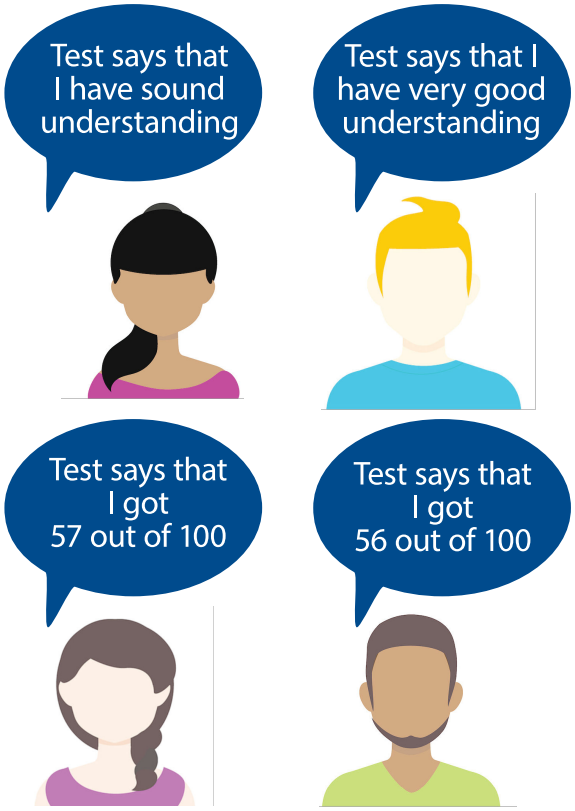
> Clearly, if the other criteria are less reliable than the examinations, greater reliance on them will lead to less reliable selection decisions.
>
> (Cresswell, 1986, p. 42)

If selection is going to take place, the IB has a responsibility to support its students by making the selection decision as fair and meaningful as possible. If the IB only provides descriptive accounts of students to the receiving institutions, they will need to find some other way of comparing students—which will almost certainly be less reliable and comparable than that offered by grading examination outcomes.

This is not to suggest that summative assessments on their own are a perfect or even particularly good way of making such a selection—but they are fairer than many of the alternatives. Furthermore, when used with other relevant indicators (such as grade point averages, extracurricular activities, personal essays and letters of recommendation), they can provide useful information for decision-making. And perhaps most importantly, the IB is constantly striving to make them as fair, meaningful and reliable as possible.

Figure 15

Should we differentiate between the two students in each pair?



In the example in figure 15, the first pair of students are distinguished using the generic grade 5 and grade 6 DP grade descriptors. If necessary, it would probably be fair to make a decision between the two students, and it is likely that if they took the tests again they would get the same outcome. In the second pair of students, however, the difference is likely not to be meaningful, and the two students probably have the same ability.

What this example demonstrates is that not all differences are meaningful, and the use of grades provides an indication of where the IB feels it can differentiate between two students. It is certainly true that for a student on the boundary (either just above or just below it), either grade is fair, but for most students a difference in grades represents a meaningful difference in overall performance.

This leads to the question of how many grades to have. If there are only two grades (pass or fail) then most students will get a fair grade, but for those on the boundary the consequences are very serious. In contrast, if there are 20 grades, then far more students lie on or close to a boundary, but the consequence of being in the wrong category is less significant for each student. Cresswell (1986) explores this concept in more detail.

The IB has generally selected **seven grades** as representing the number of meaningful categories its assessments can provide, as well as the right balance between the number of students on a boundary and the impact of being in the wrong grade.

The importance of professional judgement

The complex, higher-level thinking skills that form the focus of IB assessment do not lend themselves readily to simple, judgement-free marking. Student responses are likely to be highly varied, with many equally valid and correct forms of response. Research suggests that complex knowledge and skills should

not be taught by breaking them down into small, discrete building blocks—and the same principle applies to marking such responses. When developing markschemes, therefore, IB subject-matter experts need to provide strong guidance on marking all students in a consistent way, and how to achieve this.

This means the IB needs to place a great deal of emphasis on the professional judgement of examiners, and particularly on the professional expertise of the senior examiners who set and explain the marking standards. Taken together, this represents a strong challenge to the reliability of the assessment system, but a challenge that must be met, in the context of the precision of the outcomes.

Marking IB assessments

- Marking is the process of evaluating how well a student has completed the task they have been set.
- Marking can focus on the individual details of an answer or make a more holistic, global judgement.
- For formative assessment, marking may not result in a numerical score but instead be purely descriptive.

Defining marking

For this section, "marking" and "marking methods" refer to the IB-validated marking of summative assessments, in the MYP, DP and CP. Formative assessment is used across all programmes, but in many cases this is completed in the classroom and the IB does not validate the assessments through examiner marking.

Marking does not describe how well the student has done. Their achievement level depends on a number of things, such as how hard the questions were and what the expected pass mark is for each level. What marking does is to compare the student's answer against what an expected perfect answer looks like.

Marking can be undertaken in various ways, depending on the nature of the task and the type of marking that the assessment instrument determines to be appropriate.

Marking methods

The IB uses a range of assessment instruments, and the demands of the assessment instrument help to determine which type of marking is appropriate.

Sometimes, marking is very **objective**—the student was either correct or not. This is often the case if the answer requires only a couple of key words, or for the student to select from a number of possible answers. Automarking is the process of using technology to assess students' work against a pre-defined markscheme where an answer is objectively right or wrong. The IB uses automarking for multiple-choice questions where the answers are definitively right or wrong, and don't require an examiner's judgement to be confident that an answer is 100% correct. The IB does not utilize artificial intelligence-supported marking in place of humans where there is a narrow range of expected answers but where a judgement is required.

On other occasions, marking is far more **subjective**. An examiner might be required to judge holistically whether the student has produced an acceptable response, or which of several statements, known as "markbands", best describes a given answer's fit to the "perfect" answer. An example of the type of markbands that the IB uses is illustrated in table 5.

Table 5

An example of IB markbands

Marks	Level descriptor
0	The work does not reach a standard described by the descriptors below.
1–3	• Little understanding of the issue is demonstrated. Subject-specific vocabulary is not used, or is consistently used inappropriately. • The explanation of the issue is minimal. Points made are superficial and frequently unclear.

Marks	Level descriptor
	• The response is descriptive. Any analysis present is superficial or incoherent. There is little or no discussion of different points of view. Where a conclusion is included, this is very superficial or is not consistent with the rest of the response.
4–6	• A basic understanding of the issue is demonstrated. Subject-specific vocabulary is used, but often inappropriately. • The explanation of the issue is basic and underdeveloped. Points are often imprecise or vague, and it is often unclear what the response is trying to convey. • There is limited analysis present, and overall the response is more descriptive than analytical. There is limited discussion of different points of view. A simplistic conclusion is included.
7–9	• Some understanding of the issue is demonstrated. Subject-specific vocabulary is used, sometimes appropriately. • There is a satisfactory explanation of the issue, although the explanation lacks clarity and development in places. Relevant points are made but lack accuracy and detail. • The response contains analysis, although this analysis lacks development. There is some discussion of different points of view. A conclusion is included.
10–12	• Good understanding of the issue is demonstrated. Subject-specific vocabulary is used, mostly appropriately. • The explanation of the issue is clear but is in need of further development. Points made are relevant and accurate but lack detail. • The response contains critical analysis, although this analysis lacks development. There is discussion of different points of view. The response argues to a conclusion that is consistent with the arguments presented.
13–15	• Very good understanding of the issue is demonstrated. There is accurate and precise use of subject-specific vocabulary. • The explanation of the issue is clear and well developed. Points are relevant, accurate and detailed. • The response contains well-developed critical analysis. There is critical discussion of different points of view. The response argues to a reasoned and clearly stated conclusion that is consistent with the arguments presented.

Another consideration is whether the marking is carried out separately for several different aspects of the work, often called **criteria**. For example, an essay could be measured against four separate criteria: quality of grammar; accuracy of key facts; essay structure; and the quality of the conclusion. It is good practice with this approach to ensure that the criteria are independent of each other, as it would not be fair for a student to gain credit in several different places for the same aspect of their answer.

The opposite of **criteria marking** is **global impression judgement**. Here, the examiner internally balances all the different aspects of an ideal answer and gives a final judgement that reflects the holistic piece of work.

The major factor that can make marking hard to do reliably is the nature of the task and the breadth of the expected student responses. A broad task in which a wide range of responses are expected would be harder to mark reliably than a narrow task in which only a limited range of responses are received.

For more information on the different marking approaches, see the "Marking" section.

Marking formative assessment

All IB programmes use formative assessment. However, in the Primary Years Programme (PYP) this is the **only** type of assessment used. In formative assessment, marking does not need to be simply numeric. It is possible to compare a piece of work to the "perfect" answer and provide descriptive comments on the

similarities and differences between the two. This could take the form of shared success criteria that make clear to the student what "good" looks like. While this can make it difficult to compare how good two answers are, the aim of formative assessment is to provide feedback to support learning.

In the classroom, teachers might use formative assessment to identify gaps in a student's knowledge, understanding or skills. In this case, reliability is less important. There is no need to have a mark or grade at all. The outcome of the assessment might be a list of topics or skills that the student needs to work on. In this scenario, construct relevance is important because the student needs to be given feedback on things that are important to the course of study—or important in allowing the student to demonstrate what they know.

Nurturing an ethical mindset

Defining an ethical mindset

An ethical mindset is a commitment to act **responsibly** and **authentically**, guided by principles of **honesty**, **respect** and **integrity**. In an educational context, it fosters academic integrity, emphasizing the importance of producing legitimate and honest work. This mindset transcends rules and policies, embedding ethical decision-making and behaviour into the culture of a learning community. It encourages students to value their learning process, make principled choices and contribute to a fair academic environment.

Why create an ethical mindset?

Developing an ethical mindset is essential for ensuring the credibility and validity of educational achievements. It protects the integrity of assessments, enabling students' achievements to reflect their genuine abilities. Moreover, fostering ethical behaviour prepares students for life beyond the classroom, instilling values critical for success in higher education and professional environments. When ethical principles are ignored, trust within the school community erodes and the credibility of the institution is jeopardized. By nurturing an ethical mindset, schools empower students to become principled global citizens, committed to fairness and respect in all facets of life.

How is an ethical mindset created?

Developing an ethical mindset involves a holistic, community-wide approach. Schools should embed academic integrity into their culture, beginning with clearly communicated expectations tailored to students' developmental stages. Teachers play a pivotal role by modelling ethical behaviour, designing preventative strategies and integrating discussions of integrity into curriculums. Coordinators and school leaders must ensure policies are well defined, consistently applied and supported by training for all stakeholders. Open communication with parents and legal guardians helps align expectations and reinforces the importance of ethical practices at home.

Regular workshops, clear examples of ethical challenges, and fostering dialogue around real-world applications of integrity can further solidify these principles. By emphasizing the benefits of honesty, rather than focusing solely on penalties, schools create a supportive environment where students develop the skills and understanding necessary for ethical decision-making.

An ethical mindset is not merely a goal but a continuous practice, achieved through shared responsibility, consistent messaging and unwavering commitment from the entire school community. For further information, refer to the printable resource "Building an ethical mindset".

Ethical use of artificial intelligence in assessment

There are two main points to consider when discussing how artificial intelligence (AI) interacts with assessment.

- What impact does AI have on what is assessed?
- What role does AI have in the process of making judgements about students?

AI is now part of the world, and the responsible thing to do is to engage with it and help IB students to be informed, ethical users of this technology.

"What does it mean to say that this is my work?"

This is the key question that underpins the first challenge of what impact AI has on what is assessed. One initial fear about AI in education was that students would use it to "cheat", in the sense of asking AI to create their work for them. Various claims have also been made about how AI can do better than students on certain assessments. These concerns lead back to the most fundamental questions of assessment: what is the purpose of a specific question? What evidence about the student are we trying to gather? Facts such as the dates of a historical event can be looked up in a few seconds using the internet, but it is not always appropriate to search the internet for an answer, for example, in a time-critical situation such as landing a plane.

The same is true of other uses of AI. Sometimes it is the content and not the presentation that is critical. In this case, marks for communication style may not be necessary, and AI could support the student in presenting their ideas in a readable form. In contrast, when writing a poem, the choice of each word is critical. Students should not use AI to compose any part of a final poem, but they could use it, for example, as a thesaurus.

The critical educational question raised by the prevalence of AI is: what do students need to learn so that they can use AI responsibly and ethically in their daily lives? This key question will doubtless evolve over time. In the meantime, its answer includes educating students to:

- be critical receivers of information and be aware of potential bias or misleading conclusions
- follow up on sources and references to make sure that they are being accurately used
- understand what is—and is not—their work, and to present this honestly.

"Hello, I will be your AI assessment assistant for this examination"

AI will gradually change the way in which students can be assessed. The progression to digital assessment allows for far more interactive tasks, and the incorporation of AI can develop this further—students might, for example, explain ideas to, and interact with, digital avatars. This technology may well create opportunities for different approaches to assessment, such as making greater use of interviews, which are currently unmanageable owing to the need for human examiners with whom to interact.

Automatic marking of objective answers (such as multiple-choice or drag-and-drop questions) has been a tool used in assessment for a long time, but developments in machine learning and AI also broaden the potential for automatic marking to be used for more subjective questions and complex answers. For the foreseeable future, any final high-stakes mark will need to be determined by a human examiner, but AI offers great potential to support marking quality and identify work that should be reviewed by a second examiner.

AI marking could also support formative assessment as part of classroom practice, and it could lead to adaptive assessment opportunities in which questions are tailored to a student's ability based on their previous answers. This approach has been used for multiple-choice questions in the past, but advances in AI offer this possibility for wider question types, leading to more personalized examinations for students.

When considering the impact of any new technology, the most important factors are to ensure that the purpose of the assessment remains relevant for the intended learning outcomes, and that the assessments themselves remain valid to that purpose.

Remember, a student using AI to **support their learning** can be beneficial, while using AI **instead of learning** or to submit work that is not their own is not acceptable.

Further reading

- *Key points to explain the use of Artificial Intelligence (AI) tools (one page fact sheet)*
- *Evaluating 13 scenarios of Artificial Intelligence (AI) in student coursework (long version)*

Defining practices

- The practices described in this section apply to summative assessments in the Middle Years Programme (MYP), Diploma Programme (DP) and Career-related Programme (CP).
- They do not apply to the Primary Years Programme (PYP), where the International Baccalaureate (IB) does not provide summative assessment, or to any formative assessment the school may undertake.

This section outlines the practices by which the IB produces outcomes for those students who enter its externally marked or moderated assessments. It does not cover any other assessment that is not conducted by the IB, such as the teacher-constructed assessment that may be part of the PYP.

A **principle** sets out **why** we do something, and a **practice** describes **how** we do it. This section explains the high-level practices that the IB uses to make sure its assessment outcomes are valid.

The next level of detail is the IB's **procedures**, which describe the individual steps in delivering each practice. Where these processes relate to schools, details can be found in the programme-related *Assessment procedures* publication.

What IB assessments measure and the role of prior learning

The intention of summative assessments is to measure students' performance in terms of **knowledge**, **understanding** and **skills** as they complete their studies in the MYP, DP and CP. This has several implications, including the following.

- Summative assessment results reflect how a student is performing at a moment in time. Summative assessment does not measure a student's potential or their progress in learning.
- IB assessments are designed to minimize assessment inaccuracies that may be caused by a student underperforming on a particular question or day. This is usually achieved by having multiple examinations to give students several chances to show what they can do. However, the number of examinations must be manageable and not place an excessive burden on students.
- Summative assessment reflects the knowledge, understanding and skills of individual students and not a group of students.
- The IB does not take prior learning into account when allocating assessment grades. This means that the IB does not consider any qualifications, grades or achievements that the student obtained before they started the relevant IB programme. IB subject guides describe the relevance of prior learning to a subject or area of learning.

Reporting student achievement

- The focus of the IB mission statement is to enable the development of young people who can create a better world. Successful assessment should support this aim.
- IB grades have meaning, and grade boundaries are set taking this meaning into account.
- While grades represent a very simplified view of students' achievements, they also allow stakeholders —such as universities, employers, or colleges and schools—to make reasonable judgements around selection.
- If only more complex and holistic information is provided, then the onus is on others to simplify that evidence to make meaningful selection decisions, and they may take less care in doing this than the IB does in setting grade boundaries.
- Student achievement is more than just examination results, and even when dealing with grades, a disappointing result for one student could be a great achievement for another.

> The International Baccalaureate aims to develop inquiring, knowledgeable and caring young people who help to create a better and more peaceful world through intercultural understanding and respect.
>
> (IB mission statement, 2024)

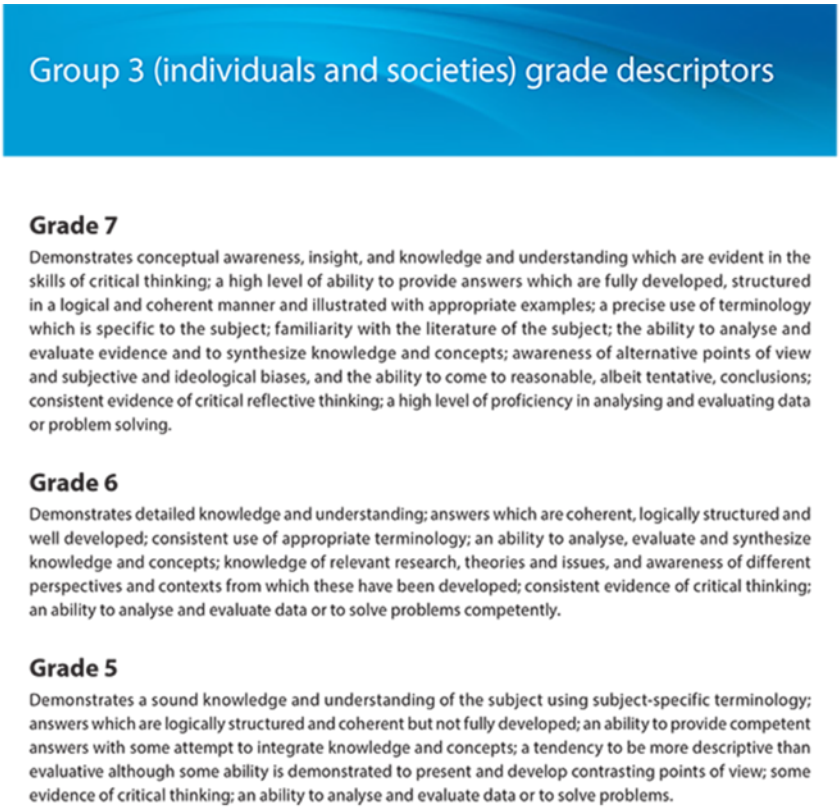
The meaning of IB grades

The outcomes of IB assessments are grades. These grades represent the standard of work that a student has shown in their answers.

The IB publishes descriptions of each grade. These descriptions are different for the MYP, DP and CP as they reflect the standard of work the IB expects from students of different ages.

Figure 16

Examples of DP grade descriptors



Group 3 (individuals and societies) grade descriptors

Grade 7

Demonstrates conceptual awareness, insight, and knowledge and understanding which are evident in the skills of critical thinking; a high level of ability to provide answers which are fully developed, structured in a logical and coherent manner and illustrated with appropriate examples; a precise use of terminology which is specific to the subject; familiarity with the literature of the subject; the ability to analyse and evaluate evidence and to synthesize knowledge and concepts; awareness of alternative points of view and subjective and ideological biases, and the ability to come to reasonable, albeit tentative, conclusions; consistent evidence of critical reflective thinking; a high level of proficiency in analysing and evaluating data or problem solving.

Grade 6

Demonstrates detailed knowledge and understanding; answers which are coherent, logically structured and well developed; consistent use of appropriate terminology; an ability to analyse, evaluate and synthesize knowledge and concepts; knowledge of relevant research, theories and issues, and awareness of different perspectives and contexts from which these have been developed; consistent evidence of critical thinking; an ability to analyse and evaluate data or to solve problems competently.

Grade 5

Demonstrates a sound knowledge and understanding of the subject using subject-specific terminology; answers which are logically structured and coherent but not fully developed; an ability to provide competent answers with some attempt to integrate knowledge and concepts; a tendency to be more descriptive than evaluative although some ability is demonstrated to present and develop contrasting points of view; some evidence of critical thinking; an ability to analyse and evaluate data or to solve problems.

While these generic grade descriptors should be the same for all subjects in a subject group, the IB often puts them in subject-specific contexts to make it easier to understand what they mean in each case. It is important to understand that the standard is not changed by this subject context. A grade 4 in a language should mean the same thing as a grade 4 in a science or a grade 4 in an arts subject. This is inherent in the IB approach to programmes, where all grades count equally. However, there has been much discussion among educationalists on whether such a concept makes sense—how can you compare achievement in two different subjects? Is it even meaningful to try? This concept is explored more fully in the section "Comparability".

Figure 17

How can you compare these two pieces of work?

9. (a) $x = e^{3y+1}$

taking the natural logarithm of both sides

$$\left(f^{-1}(x) \right) = \frac{1}{3} (\ln x - 1)$$

(b) coordinates of Q are $(1,0)$

$$\frac{dy}{dx} = \frac{1}{x}$$

at Q, $\frac{dy}{dx}$

$y=x-1$

(c) let the required area be *A*

$$A=\int_1^e 1dx-\int_1^e \ln x dx$$

use integration by parts to find $\int \ln x dx$

$$= \left[\frac{x^2}{2} - x \right]_1^e - \left[x\ln x - x \right]_1^e$$

$$= \frac{e^2}{2} - e - \frac{1}{2} \left(\frac{e^2 - 2e - 1}{2} \right)$$

As part of validity considerations, the argument for comparability between IB grades in different subjects rests on the purpose of IB grades. IB grades are intended to allow stakeholders to compare students' attainment. Thus, it is meaningful to use statistical and qualitative methods to aim for parity in the meaning of grades.

The difference between marks and grades

Marks and grades are not the same thing.

Figure 18

A metaphor to compare marks and grades

We would expect "good" candidates to have completed most of the task.

We would expect "good" candidates to have only completed part of the task.

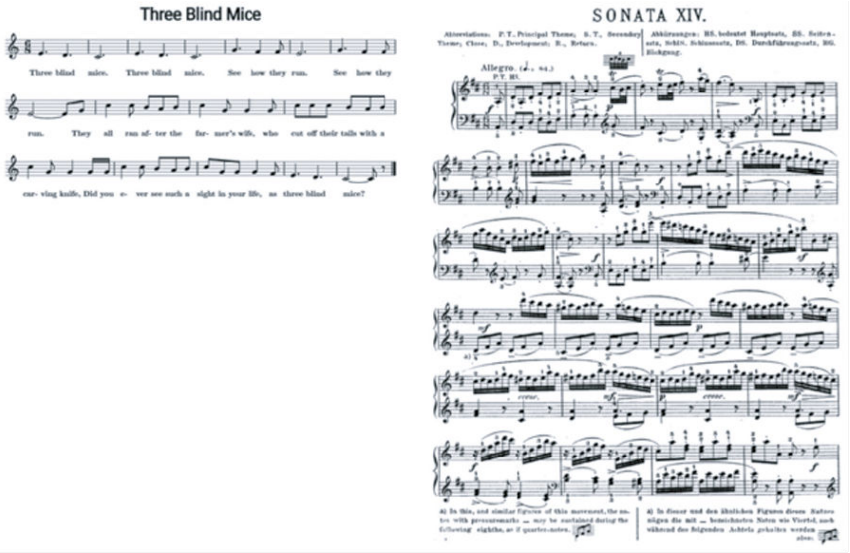
There are many metaphors available to explain the difference between marks and grades. For example, in figure 18, the distance walked could represent marks, as a common measure of how far someone has travelled. But to understand how much of an achievement it is to walk any given distance, it is necessary to consider where a person was walking—this is taken into account in setting grades.

- **Marks** represent **how much** of a task a student has completed.
- A **grade** takes into account **how difficult** a task is, in order to quantify the achievement level represented by the student's mark.

Consider the two examples provided in figure 19. In the first example, a 16-year-old would be expected to get nearly all of the task right to get a "good" grade. In the second example, far less would be expected to indicate the student deserved a "good" grade.

Figure 19

Assessment tasks with varying levels of difficulty



We would expect "good" candidates to have completed most of the task.

We would expect "good" candidates to have only completed part of the task.

This consideration leads to one of the challenges of setting quality assessments. Students need the opportunity to show their full potential, which may not be possible if tasks are too simple. Conversely, if tasks are too challenging, this may produce a situation where too many students cannot even start the task. This makes it difficult to differentiate between them using grades.

Overly simple tasks may also lead to the measuring of accuracy rather than understanding; as a result, students who make minor mistakes may be excluded from the top grades despite having a good comprehension of the topics.

For more information on the difference between marks and grades, see "The difference between marks and grades: How we take into account the difficulty of an assessment, and not just how much a student got right" (video).

A successful examination session

Figure 20

Differing perspectives on a successful examination session

What success looks like depends very much on one's point of view. The IB focuses on both high-level construct relevance and practical delivery. To be considered successful, an examination session would result in the following.

- The assessments allowed every student a fair opportunity to show their ability.
- The experience for schools and students was as straightforward as possible.
- All stakeholders retained confidence in the outcomes (grades) that the IB released.

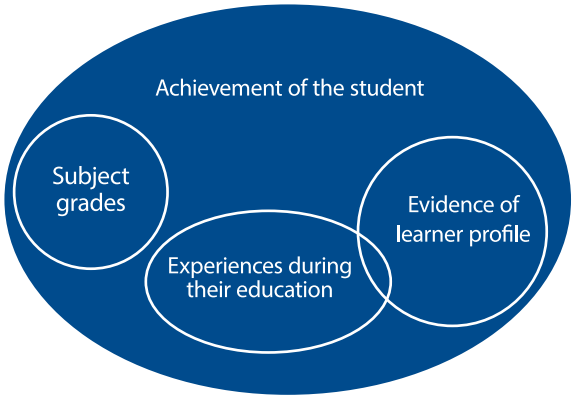
Grades and achievement

As shown in the IB mission statement, the goal of an IB education is far more than a series of academic grades. This is reflected in the IB learner profile, and articulated in *What is an IB education?* (2019).

Assessment outcomes can only focus on a narrow part of this mission statement. Given the compensation model used in IB examinations, what proportion of marks in a mathematics assessment should be allocated to caring or risk-taking, even assuming that giving a numerical value to these important traits is reasonable? One of the principles of IB assessment is that it should have a positive backwash effect, which would support these goals.

Figure 21

A holistic picture of student achievement



When reporting the outcomes of an IB programme, it is important to consider more than just assessment grades in order to reflect the full range of student achievement.

The IB only records the final attainment of students in the assessments, without any indication of how challenging it was for them to achieve this result, or their full potential. The IB recognizes the importance of both these elements, but believes it is not possible to measure them meaningfully within summative assessment. Such evaluation is the responsibility of the school, which has a holistic view of the student.

While there are methods to calculate "value-added" measures of student attainment or "predicted grades based on prior attainment", caution is advised regarding their use as an indication of student success. Such measures are based on an average student, and each student is a unique combination of personal characteristics and traits that should be celebrated appropriately by those who have the opportunity to learn with them over a whole programme.

Predicted grades

As stated in the *Programme standards and practices* (2020) publication, "The school communicates … IB predicted grades to the IB as part of IB-validated assessment, in ways that value academic integrity and reflect as much accuracy as possible".

Given that student performance can vary throughout the course of study, predicting final grades can be challenging. Predicted grades serve several important functions. They provide evidence during grade award meetings, which help the IB understand the overall strength of the cohort and the grade distributions for each subject. They also help identify large discrepancies between predicted and actual grades, enabling additional quality checks to be conducted, if necessary. For schools, predicted grades help set and manage the expectations of students and their parents and legal guardians.

While it's unlikely that a teacher will predict all their students' grades with complete accuracy, the aim is to be as precise as possible. The more accurate the predicted grades, the more useful and constructive they become within the process.

A predicted grade is an important data point for assessment; it is taken into consideration at an overall subject level to set subject grade boundaries, and at a school level provided there is evidence of accuracy. Therefore, it is important that predicted grades are accurate so that they can be treated as reliable data points.

A predicted grade is also:

- a data point for schools, who can choose whether to share them with students or not
- an indication of capability for post-programme options, setting students up for success
- a grade (1–7; A–E) as opposed to a mark.

When predicting grades, it is important to understand what a predicted grade is not. A predicted grade is not:

- a mechanism to compensate for adverse circumstances
- a disciplinary tool
- an equation based on a previous session's grade boundaries
- to be confused with other uses of predicted grades, such as for university or college entry
- a mechanism to appease students and their parents and legal guardians
- an exaggeration of capabilities in either direction.

For more information on predicted grades and for strategies to improve accuracy, see the printable resource "Predicted grades: A teacher's guide" (PDF) and additional guidance on the PRC as follows.

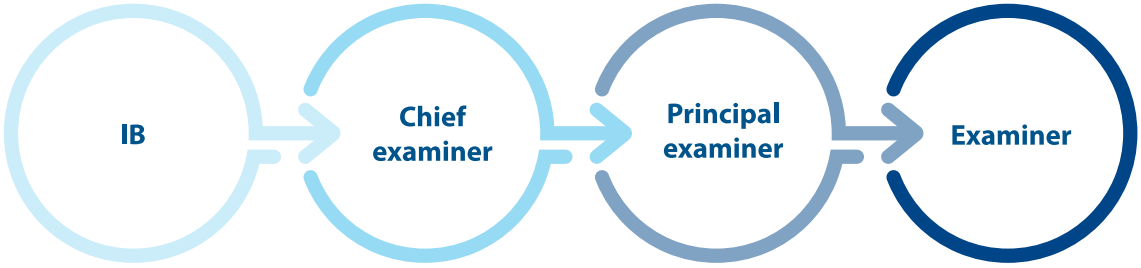
- **MYP resources > MYP eAssessment > Cross-session resources >** ***Predicting IB Grades***
- **DP resources > Assessment > Cross-session resources >** ***Predicting IB Grades***
- **CP resources > Assessment > Cross-session resources >** ***Predicting IB Grades***

Assessment process: Roles and responsibilities

- Each of the roles in the assessment process has its own responsibilities and skill set.
- In some cases, the same person may fulfil different roles at different points of the assessment cycle.
- Some of these roles are fulfilled by the IB, while others draw upon experts from the IB community. In the latter case, the IB is responsible for the final sign-off, and for maintaining the quality of these aspects of the assessments.

Figure 22

Accountabilities and responsibilities of the key roles in the assessment cycle



IB
- Accountable for all aspects of assessment
- Responsible for assessment processes such as examiner recruitment, examiner quality and issue of results
- Makes decisions on issues relating to student academic misconduct, school maladministration, inclusive access arrangements and adverse circumstances
- Accepts or challenges the grade boundaries recommended by the chief examiner

Chief examiner
- Takes overview of all components in their subject (group)
- Ensures consistency of standards between all components, including in examination authoring
- Arbitrates on any academic issues relating to the assessment
- Recommends final grade boundaries to the IB

Principal examiner
- Responsible for one component
- Final arbiter on what mark students' answers receive in that component
- Ensures that all examiners understand the marking standard in that component
- Guides chief examiner in setting grade boundaries for the component

Examiner
- Marks student work in accordance with the standard set out by the principal examiner

Principal examiner and chief examiner

The principal examiner:

- manages a specific component, such as internal assessment (IA)
- sets the marking standards in a session and between sessions, and explains them to examiners through standardization discussions and quality models

- guides the chief examiner on grade boundary recommendations for the component for which they are responsible
- is usually an assessment content developer, unless there is a conflict of interest
- is engaged by the IB for each examination session as an external education professional and usually contributes across multiple sessions.

The chief examiner is responsible for maintaining the quality of several related components. In the DP and CP, they are responsible for a whole subject; in the MYP (which usually only has one component per discipline), the chief examiner is responsible for a subject group, such as sciences or language acquisition. They act as the IB's academic expert in this field.

The chief examiner:

- ensures quality across related components for a whole subject (DP/CP) or subject group (MYP)
- acts as the IB's academic expert, maintains standards and ensures consistency
- assists in resolving disputes
- leads the grade award process and recommends grade boundaries to the IB
- often also acts as a principal examiner for a given component
- is engaged by the IB for a specific period.

For small-entry subjects, the IB appoints examiners to the role of examiners responsible. Examiners responsible hold the same duties as chief examiners relative to their subject, which usually comprises fewer principal examiners and examiners. Chief examiners are also invited to work with the IB in discussing and improving its curriculums and assessments through curriculum review activities, which are not discussed in this publication.

Other examiner roles

Examiners are responsible for marking work submitted for assessment to the standard set out by the principal examiner. They need to prove they have understood this standard and are applying it through the quality model.

Examiners apply to the IB to mark in a session. They must be experts in the subject they wish to mark, usually by being a teacher of the subject. An examiner must also have experience teaching students in the relevant age range. The IB verifies credentials through a rigorous application process, which includes the provision of referees. Examiners are typically only offered one coursework component and one examination component to mark per session. The marking period for these two components will not overlap. Examiners are paid per "live" student script. Live scripts do not include any qualification or seed scripts (see the "Quality model" section) which simply demonstrate to the IB that the examiners understand the marking standard.

Team leaders are particularly experienced examiners whom the IB asks to support other examiners in understanding the correct marking standard. Team leaders support examiners through the qualification process and provide feedback if their marking of seed scripts indicates that they are too far from the definitive standard. The number of team leaders in a subject will depend on the number of examiners overall, which depends on the cohort size. In a small-entry subject, the principal examiner may provide this support to all examiners.

Standardization team members are usually experienced examiners who are part of the qualitative discussions that lead to the development of a quality model. Standardization team members discuss the application of the principal examiner's standard to ensure that it can be applied and adhered to.

For more information on the role of examiners, see "Introducing: The examiners: Who are they and what do they do?" (video).

The responsibility of IB staff

Subject managers and assessment operations analysts are among the roles involved in ensuring reliable assessment from an IB perspective.

Separate teams within the IB manage requests for modified papers and other inclusive access arrangement and adverse circumstances requests, issues related to academic integrity, and the monitoring of examiner quality and school moderation factors.

The director of assessment, supported by the head of assessment development and delivery and head of assessment principles and practices, is responsible for considering the recommended grade boundaries put forward by the chief examiner as a culmination of a grade award meeting for each subject, and either approving them or asking the chief examiner to reconsider the recommendation.

Ultimately, the IB is accountable for all decisions made as part of the assessment cycle. The use of external experts supports the IB in producing fair, high-quality assessments, but the final accountability rests with the IB.

Roles in authoring examinations

Assessment content developers have a designated role within the examination authoring process, which can include the work of an author, reviser or scrutineer. The key skills required for assessment content developers include the following.

- Creativity: crafting engaging, curriculum-aligned questions
- Cultural awareness: ensuring questions are unbiased
- Communication skills: writing clear, unambiguous questions that translate well and that test the intended skills, leading to reliable marking.

Other roles in assessment production include but are not limited to the following.

- Technical design editors: formatting questions for examinations
- Translators: converting assessment into required languages
- External reviewers: providing feedback on length and clarity by taking the assessment as students

For more details on the way that assessments are authored, see the section "Assessment production".

Examiner hierarchy

- The principal examiner is the final arbiter of what mark to award on a component, and their standard is the definitive standard. All other examiners must align with the definitive standard.
- The IB makes the assumption that those who have been involved in setting the standard have a better understanding of the principal examiner's standard than those who have learned it from the quality model and their team leaders.
- "Senior" refers to closeness to the principal examiner in the examiner hierarchy. It does not refer to the length of time that a given person has been an examiner, nor any associated teaching experience.

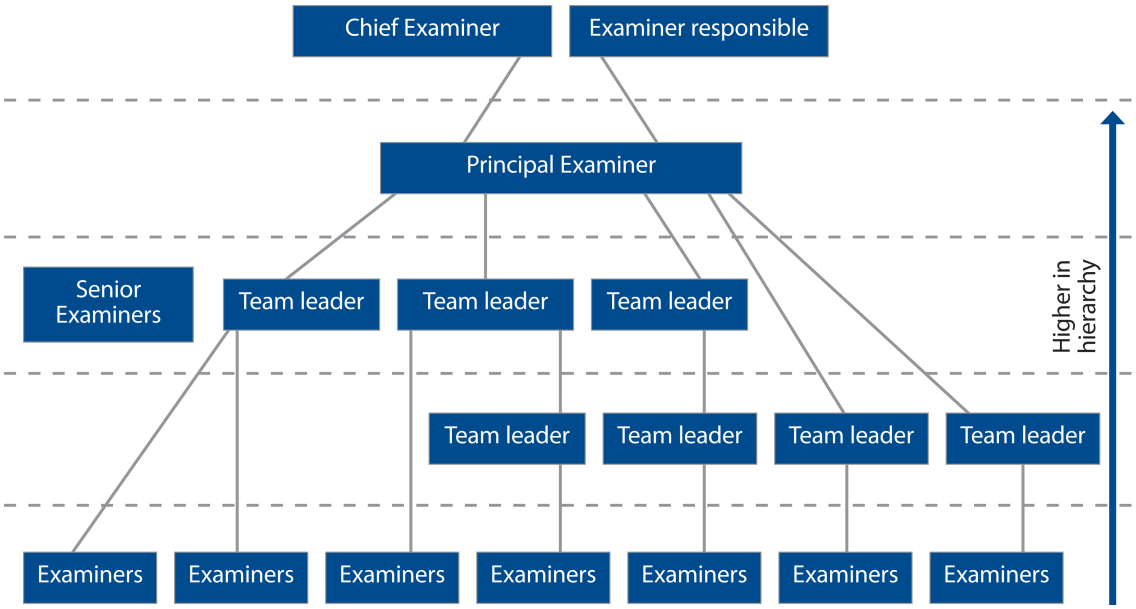
The principle behind the IB's marking is that the principal examiner is the final arbiter of what the correct mark for a student's work is. While the IB recognizes that there may be many different and equally reasonable views on what the "correct" mark is, it is not fair for a student to get a different mark depending on which examiner reviewed their work. Therefore, the principal examiner sets the standard and asks every other examiner to follow it.

The principal examiner is supported in setting the standard by a team of senior examiners. The IB makes the assumption that, as these examiners have had the opportunity to discuss the standard with the principal examiner, they will have the best understanding of it. However, marking reliability of all examiners aligned with the principal examiner's standard is plotted each session to give an indication of examiner reliability.

There is a hierarchy of examiners. For a particularly large subject, there might be a structure like the example in figure 23.

Figure 23

Hierarchy of examiner roles for a very large subject



The hierarchy can be flexible. If, as a result of data from qualification or seed scripts, it is shown that a given examiner has a better grasp of the principal examiner's standard than a more senior member of the hierarchy, then this will be taken into account. A team leader must be able to explain the standard and not just to apply it, so reliability is not the only consideration in designating team leaders.

Ideally, there would never be a need to make decisions based on the hierarchy as every examiner would mark to the same standard. In reality, the hierarchy is an essential way of determining which marks the IB should use. The IB would refer the script for a particularly difficult or controversial case to the principal examiner to determine the appropriate mark.

Assessment integrity

- IB assessments can only be fair if all students have an equal opportunity to achieve.
- School maladministration and student actions that contravene the academic integrity policy create a disadvantage for those students who have followed the rules. The IB takes every effort to prevent such behaviour.
- The IB rules, set out in the "General regulations" section of *Assessment procedures* for each programme and in the *Academic integrity policy*, are designed to minimize the opportunity for student academic misconduct or school maladministration. Ultimately, only the school can create the learning culture where academic misconduct is not acceptable and is reported.
- While certain forms of assessment are less susceptible to academic misconduct than others, the IB's principle remains that construct relevance (that is, testing what we really want to assess) should remain its primary consideration when designing assessments.
- Any suspicion of student academic misconduct or school maladministration must be reported to the IB.

Defining academic integrity

Academic integrity is a guiding principle in education and a choice to act in a responsible way whereby others can have trust in us as individuals. It is the foundation for ethical decision-making and behaviour in learning, teaching and assessment, and in the production of legitimate, authentic and honest scholarly work. It should not be imposed as a series of strict rules, but should instead be a positive culture within a school and the wider community, including parents and legal guardians, which organically leads to fairer assessment outcomes for all.

In order for assessments to be valid, they need to provide an accurate reflection of a student's achievement relative to all the other students who have taken the assessment. For this reason, the IB takes great care in having consistent approaches to marking, grading and removing bias from its examinations. The rules and regulations it sets out are another aspect of creating this equality of opportunity and ensuring fairness.

The *Academic integrity policy* defines academic misconduct as deliberate or inadvertent behaviour that has the potential to result in the student, or anyone else, gaining an unfair advantage in one or more components of assessment. Such activity affects not only the students involved, but everyone who has taken the assessment, as it reduces the validity of the qualification. The IB therefore takes academic misconduct very seriously, and details of its prevention and consequences can be found in the *Academic integrity policy*.

Assessment design is an important tool in preventing academic misconduct as some types of assessments are easier to monitor than others. For example, in an examination, it is harder to obtain help from someone else than it is for an internally assessed piece of work. While the IB takes such considerations into account in devising its approach to assessment, the principle is that it should not sacrifice construct relevance to prevent the opportunity for academic misconduct.

Manageability is another important aspect of validity, and this must also be considered when preventing the opportunity for academic misconduct. Setting up examination rooms, particularly with digital assessment, can be a real challenge for schools, and this must be taken into account. The IB endeavours to keep the experience of schools in mind when setting its rules and regulations, and is keen to hear from coordinators and head teachers about positive and challenging practices.

For more information on academic integrity, and what it means to act ethically, see "Academic Integrity in the IB: Making the Right Choices" (video).

Conflicts of interest

Within the IB, access to examination content is carefully controlled and the IB actively manages any connections to students taking its examinations that could constitute a conflict of interest. Staff responsibilities are reorganized if such a conflict could be seen to occur.

The IB principle is that no examiners can mark their own students' work, or the work of students from schools with which they have a conflict of interest (for example, a school where the examiner tutors another subject, or a school that the examiner has recently left). On the extremely rare occasions when this is unavoidable because of circumstances beyond the IB's control (for example, if there is no one else qualified to mark at the required standard), a second independent examiner would review the marking. The IB would need to establish that there was no indication that a different standard had been applied to the examiner's own students.

Managing school maladministration and student misconduct

For each programme, the "General regulations" section of the relevant *Assessment procedures* publication and the *Conduct of examinations booklet* or *The Conduct of IB Middle Years Programme on-screen examinations* set out rules and instructions that minimize the chance of maladministration occurring. For paper-based examinations, the *Secure storage of confidential IB examinations material booklet* provides instructions on the storage of confidential IB materials on school premises and the measures that schools need to take to implement the policy.

Examinations must be invigilated according to the instructions provided in the *Conduct of examinations booklet* (DP and CP) or *The conduct of IB Middle Years Programem on-screen examinations* (MYP). Invigilators should remain vigilant throughout an examination to ensure no academic misconduct takes place.

The IB carries out unannounced examination visits during assessment sessions to ensure that all required examination practices are in place. These inspections are spot checks on the school's implementation of IB processes. However, the IB places great responsibility for preventing maladministration on Heads of Schools and programme coordinators, who are able to ensure high standards are maintained on a daily basis. Further, they should ensure that the culture in the school is one that encourages best practice and high levels of integrity from its teachers and students.

In addition to maintaining academic integrity standards during examinations, it is important to ensure that work undertaken in the classroom (and at home) is the student's own work, and that plagiarism does not occur.

Many websites and artificial intelligence (AI) tools exist that offer "support" to students with their work. The best defence against this kind of academic misconduct is the teacher, who will have worked with the student and can identify where the work submitted does not reflect the student's usual standard. For this reason, both the student and teacher are asked to confirm that the work is the student's own. Simply relying on plagiarism-checking software is not sufficient: teachers should also work with students as they write their IA to check the authenticity of the work. Once the work is submitted, the IB is not obliged to accept an alternative version should plagiarism be found to have occurred.

No level of plagiarism is acceptable: all citations from other authors and use of AI software must be properly referenced as set out in *Effective citing and referencing* and the *Academic integrity policy*. The IB uses a range of software to identify potential cases of student academic misconduct. Where evidence is found, a formal investigation is initiated.

Establishing and managing a culture of academic integrity is a requirement for all IB World Schools, and repeated infringements will have consequences for authorization.

Challenges with international examinations and time zones

The IB faces specific challenges around the international nature of its schools as, unlike most national systems, not all IB students can sit the examination at the same time. This means that some students will have finished their examinations before other students have started them; the IB, therefore, requires integrity in its students and teachers not to take advantage of this. Where student numbers are large enough to make it viable, there are separate "exam zones" to reduce the time between students finishing the examination in one time zone and others starting in a different time zone, but there can be considerable variation in the start time of an examination.

The IB has several measures in place to mitigate this risk, including careful monitoring of websites and social media platforms where examination content could be shared. The rule that examination content cannot be taken out of the examination room limits the opportunity to share questions online. Supervision rules that stipulate that students must remain supervised for at least two hours from the start of a morning or afternoon examination—if the examination is shorter than two hours and it is the only examination that morning or afternoon—also help reduce opportunities for the prohibited sharing of information.

The IB also ensures that examination questions are designed to test understanding rather than knowledge recall, meaning there would be limited value to knowing the questions in advance. The final mitigation is that students are required not to discuss the examination for 24 hours after it takes place, ensuring there is no possibility of "innocent discussion".

The IB continues to seek to innovate in this area to reduce the risks from time-zone cheating.

Benefits of digital assessment

Digital assessment offers many benefits for managing academic misconduct. The time in which examination content can be accessed is greatly limited, and the IB is able to monitor when examination content is accessed, and by whom.

Additionally, any modifications to the examination made by the school—such as extra time or pauses—are recorded and, where appropriate, justification for these must be provided. This means the IB can devolve more responsibility to schools for reasonable modifications while ensuring fair practice.

Digital assessment also allows for a more detailed understanding of the student's experience of the examination using metadata, which can be used as evidence to support an investigation into an alleged incident of academic misconduct. Also, the electronic nature of the digital responses allows for large-scale checking for suspiciously similar answers, which is not possible with handwritten scripts.

The IB recognizes that digital assessment can be subject to new forms of academic misconduct, particularly hacking attempts, and the IB has a range of tools and processes in place to counteract these.

Resources

Teachers, schools and students are best placed to challenge and stop cases of academic misconduct by creating a culture where it is not acceptable and being vigilant in tackling it when it occurs. This topic is explored further in the section "Nurturing an ethical mindset" of this publication. Materials to support schools in creating a culture of academic integrity, including the *Academic integrity policy*, can be found on the IB website under "Academic integrity".

If you would like further support, have any concerns about academic misconduct that is not being addressed at your school, or have thoughts about new and emerging issues in academic integrity, please contact IB Answers.

Prioritizing fairness for all

- To maintain validity, formative and summative assessment must consider learner variability and not discriminate against any student. IB assessments are designed with this learner variability in mind from the outset.
- The best way to ensure fairness is through designing assessment to be accessible for everyone. IB assessments follow the principles of universal design, eliminating barriers that could affect their validity. Schools must follow the guidance established by the IB to ensure that the conditions under which written, oral and digital examinations are conducted do not provide an advantage or disadvantage to any student.
- The *Access and inclusion policy* addresses situations where additional measures or arrangements are needed to facilitate assessment access, such as modified papers.
- In the case of unforeseen circumstances that could affect access or the validity of the assessment, a series of measures to mitigate the negative effects of such events are covered by the *Adverse circumstances policy*.
- The learning and teaching phase, along with formative assessment, should be aligned to prepare students for summative and/or IB assessments. Given that one of the key principles in providing inclusive access arrangements is that they must be used during learning, teaching and assessment (as the student's usual way of working), this alignment becomes particularly important. The alignment is also essential through all mediums of assessment, including digital assessment.
- The purpose of these arrangements is to ensure fairness for all students. In making decisions, the IB must consider what is fair for all students.

All students should be allowed to demonstrate their ability under assessment conditions that are as fair as possible. Some standard assessment conditions may disadvantage some students by preventing them from demonstrating their level of attainment. Adverse events or circumstances beyond their control may also impact the student's performance. Both factors must be considered to implement measures that ensure fairness.

The best way to ensure fairness in assessment is for all students to take the same assessment under the same conditions. To achieve this, assessments are developed with careful consideration of the diverse requirements students may have. The most effective approach to address learner variability and meet students' needs is to design assessments based on the principles of universal design for learning and universal design for assessment. Once assessments are designed following these principles, the next step is to ensure accessibility through inclusive access arrangements, as outlined in the IB *Access and inclusion policy*.

Inclusive access arrangements are designed to remove barriers and provide access to all students. These barriers might arise when assessment does not adequately consider or accommodate neurodiversity, disabilities, challenges or diverse socio-emotional circumstances.

Any reasonable adjustments for a particular student pertaining to their unique needs will be considered. For further details, please refer to the *Access and inclusion policy*.

Adverse circumstances refer to situations beyond the student's control that might negatively impact their performance. The *Adverse circumstances policy* outlines the process of addressing and mitigating these impacts. Inclusive access arrangements are put in place to eliminate barriers and ensure that all students have access to assessment. These arrangements do not provide an advantage to students facing such barriers—rather, they enable equitable access, allowing all learners to demonstrate their skills, knowledge and abilities on the same basis as other students.

Principles for inclusive access arrangements

The principles for inclusive access arrangements are set out in the *Access and inclusion policy*. Assessment must take into consideration the developmental stages of young people, reflected by the aims of each programme. Inclusive access arrangements must be considered in all programmes, both in formative and summative assessment, and throughout the learning and teaching experience.

The IB must ensure that a grade awarded to a student in any subject accurately reflects their level of attainment. To achieve this, the same standards of assessment are applied to all students, regardless of whether they require inclusive access arrangements.

Inclusive access arrangements, including reasonable adjustments, are pre-examination measures for a student to access the assessment. They cannot be requested retrospectively for either oral, written or digital examinations.

The arrangements requested for a student must not give that student an advantage in any assessment component.

The inclusive access arrangements are intended for students with the aptitude to meet all assessment requirements.

When inclusive access arrangements are necessary for a student, the school may provide them based on the IB criteria clearly set out in the *Access and inclusion policy*. The requested arrangements for assessment must reflect the student's usual way of working. It is the school's responsibility to select arrangements that best meet the student's needs throughout the course of study, ensuring a smooth transition to summative assessments. Requests for arrangements that do not align with the student's typical practices will only be approved in very exceptional and unusual cases.

The IB is committed to an educational philosophy based on international-mindedness. To achieve equity among students with assessment access requirements, the policy represents the result of a consideration of accepted practice in different countries and regions (territories).

Each request for inclusive access arrangements will be judged on its own merit. Previous authorization of arrangements, either by the IB or another awarding body, will not influence the decision on whether to authorize the arrangements that have been requested.

The IB treats all information about a student as confidential. If required, information will only be shared with appropriate IB personnel and members of the IB Final Award Committee, who will be instructed to treat such information as confidential.

If a school does not meet the conditions specified by the IB when administering inclusive access arrangements, or makes arrangements without authorization, the student may not be awarded a grade in the subject and level concerned.

If it can be demonstrated that a student's lack of proficiency in their response language(s) arises from an identified learning support requirement, inclusive access arrangements may be authorized.

Examiners must not be informed of a student's circumstances or of adverse conditions.

In the case of internally assessed work, teachers must not make any adjustments when marking a student's work.

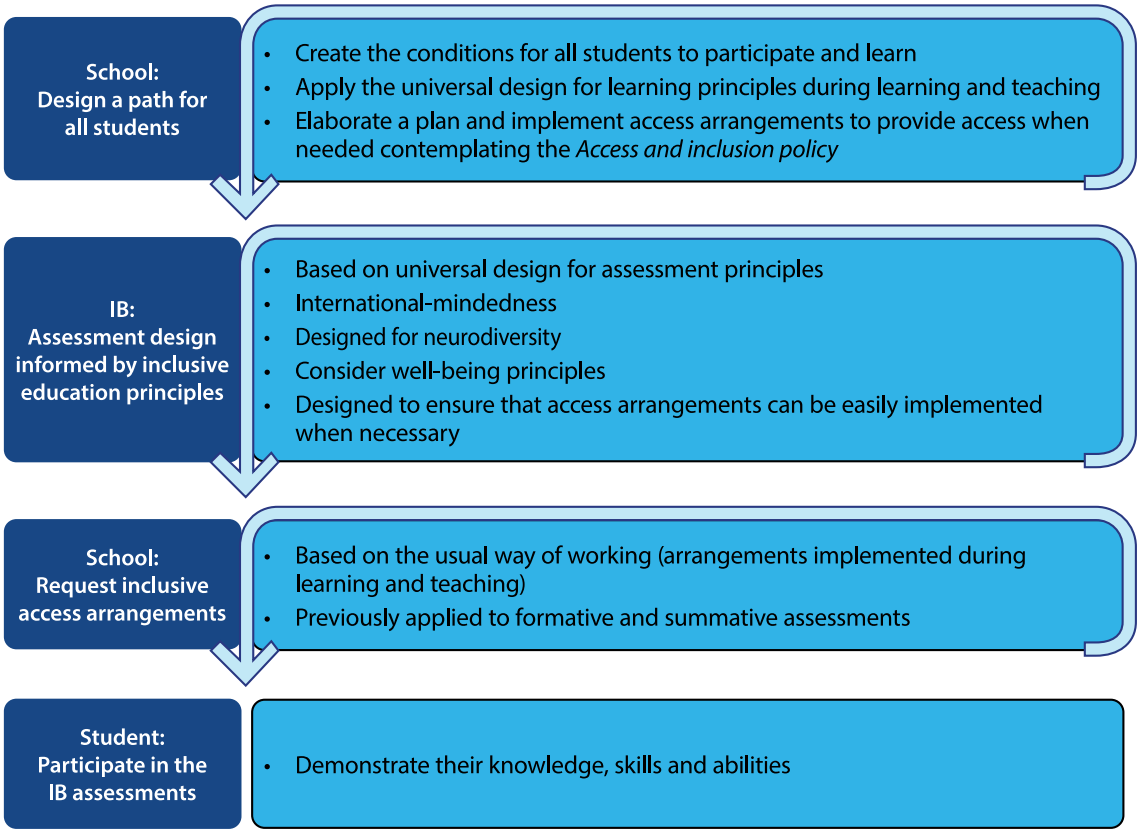
The list of inclusive access arrangements available is revised regularly. Alternative arrangements proposed by a school might be considered, provided those arrangements could be made available to all students with similar requirements.

If the student and/or the authorized inclusive access arrangement(s) might disturb other students during an examination, the student must take the examination in a separate room and be supervised according to the regulations governing the conduct of each programme's examinations.

Any issues that arise from the nature of the inclusive access arrangements, or any unforeseen difficulties encountered by the student during the examination, should be reported to IB Answers as soon as possible.

Figure 24

Implementing inclusive access arrangements



Exemptions from assessment

Exemptions are not normally granted for any assessment component. However, if an assessment component, or part of a component, demands a physiological function that a student is not able to perform, an exemption may be authorized. Before submitting a request for an exemption from a component, careful consideration should be given to whether all reasonable adjustments have been considered. Authorization for an exemption will only be given when there are substantial grounds for such an exemption. A student's inability to perform the functions required by the component must be clearly and fully documented.

For full details on the principles and processes around exemptions from assessment, please refer to the list of inclusive access arrangements in the *Access and inclusion policy*.

Opportunities for inclusive access arrangements with digital assessment

Digital assessment gives students greater control over how they wish the assessment to be presented. Computing devices offer a wide range of fonts, text sizes and colours to meet individual preferences and needs, making adjustments universally accessible to all students. While implementing universal design principles improves accessibility, some arrangements will still need to be made to provide access. To achieve this, digital features are considered to provide alternatives to implement such arrangements. Available arrangements are listed in the *Access and inclusion policy.*

The objective of inclusive access arrangements for digital examinations is to remove barriers, just as in paper-based examinations. To ensure fairness, approved access arrangements must be consistent across digital and paper-based examinations.

Adverse circumstances

Adverse or unforeseen circumstances are those that are beyond the control of the student and that might have a negative impact on their performance. This includes temporary illness or injury, severe stress, exceptionally difficult family circumstances, bereavement, or events that may threaten the health or safety of a student. Adverse circumstances may also include an event that affects the whole school community, such as civil unrest or a natural disaster.

Adverse circumstances do not include shortcomings on the part of the school. It is the school's responsibility to ensure that all students comply with programme and assessment requirements, including issues with teaching staff.

Full details of what is included and excluded within the category of adverse circumstances, and the mitigations offered, can be found in the *Adverse circumstances policy*.

Universal design for assessment

The preceding sections have discussed how situations requiring inclusive access arrangements are managed to ensure fair access to IB assessments. However, the ideal solution is to design assessments without barriers to participation in the first place. The concept of universal design for assessment emphasizes the consideration of access, inclusion, equity, cultural diversity and sensitivity. By following this approach, IB assessments are devised to ensure that many barriers are removed from the overall assessment model during the design, development and implementation process.

Creating more inclusive and less construct-irrelevant assessments from the outset considerably minimizes the further need to initiate or implement arrangements, benefiting a larger number of students.

Universal design for assessment is an aspect of the broader framework of universal design for learning, which aims to create accessible learning environments for all learners. This approach considers a broader range of diversity, including but not limited to neurodiversity, preferences, abilities, challenges, interests, cultures, languages, backgrounds.

The principles of universal design for learning are applicable across various aspects of education, including curriculum and assessment design, curriculum and assessment development, and school management. For more details on universal design for learning in the IB, refer to Rao et al. (2016).

The assessment life cycle

- The process of creating assessments should be thought of as a continuous cycle whereby each stage is informed by the previous stage and leads into the next stage.
- It takes on average 18 months to create an examination (and its corresponding markscheme), so the IB creates examinations for different sessions in parallel with each other.

The life of an assessment can be thought of as a cycle—from its creation, through sessions taken by students, marking by examiners and the results released. The critical aspect of the process is that the IB learns from each session to improve the quality of the assessment for the next session.

Figure 25 is only one way of describing the process. Many of the steps could be separated out or combined in different ways, but it is a good way of explaining the life cycle of an assessment. Table 6 summarizes each step and provides a link to the relevant section.

Figure 25

The assessment cycle

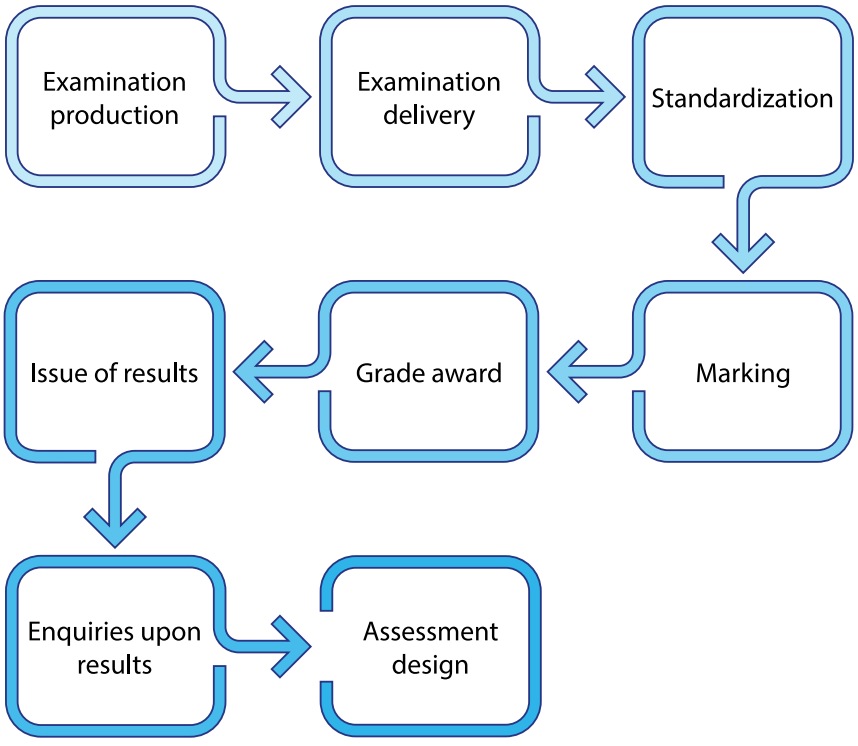


Table 6

Stages in the life cycle of an assessment

Stage of assessment cycle	Description
Examination production	The process of creating each individual examination. This covers everything from deciding on the topics the questions will cover to writing and editing the specific questions, translating them into other languages, building them in the correct format, and, finally, doing the necessary quality checks.

Stage of assessment cycle	Description
Examination delivery	This part of the process happens when the students take the assessments in schools.
Standardization	This is the process whereby the senior examiners explain how to mark students' work to their examiner teams, and identify the definitively marked scripts that will be used in the quality assurance processes for marking.
Marking	This involves examiners looking at individual students' work and deciding how many marks to award. They must follow the instructions set by the principal examiner and are regularly monitored to make sure they are doing so correctly.
Grade award	Senior examiners decide how marks (which depend on the examination) should be converted to grades (which always mean the same). Post-grade-award activities include the monitoring of examiner grade distributions and identification of students at risk of getting an incorrect grade based on the evidence of a particular issue.
Issue of results	The release of results to schools and students.
Enquiries upon results	The opportunity for schools to highlight where they think there has been an error in the examination process and ask the IB to look again at a student's work.
Assessment design	Students' performance and interpretation of examination questions informs the design of assessments. The IB uses this information to improve what is intended to be assessed and the approach the IB takes, including the number and type of tasks.

For more information on the assessment life cycle, see "The life cycle of an exam: The steps that we go through from an exam being created to grades being awarded" (video).

The impact of digital assessment on the assessment cycle

The introduction of digital assessment does not change the principle of the assessment cycle. It will allow the IB to be quicker in certain parts of the cycle (for example, it removes the need for sending paper-based examinations and scripts to scanning centres). However, each part of the assessment cycle will still need to be completed.

Assessment production

- There are several stages of finalization that an assessment must pass through during production before it can be taken by students. These include:
  - assessment authoring and content sign-off
  - building and proofreading
  - usability
  - translation
  - delivery of the assessment to schools.
- Any modifications required by students (inclusive access arrangements) are also considered during the process of finalization.
- Whether delivered as a paper-based examination or a digital examination, the principles of the assessment production cycle do not change.

Good assessment components

Alongside the qualities that make a good assessment described in "Section A: Principles of assessment", the IB expects its curriculum developers and authoring teams to keep in mind the following key elements.

- The assessment should provide opportunities for students to demonstrate what they can do, not identify what they cannot do.
- The assessment needs to be accessible to the widest possible range of student expertise, but also allow for reasonable differentiation.
- The best assessments are accessible to all. The IB utilizes universal design throughout assessment production to reduce the need, as far as possible, for inclusive access arrangements to be made.
- The assessment must test only the curriculum as set out in the subject guide.
- Consideration of how the assessment will be marked while it is being developed ensures that what is trying to be tested is what is awarded credit. This means that the markscheme is developed along with the examinations, and that guidance on the anticipated tolerances for examiners is included.

Command terms

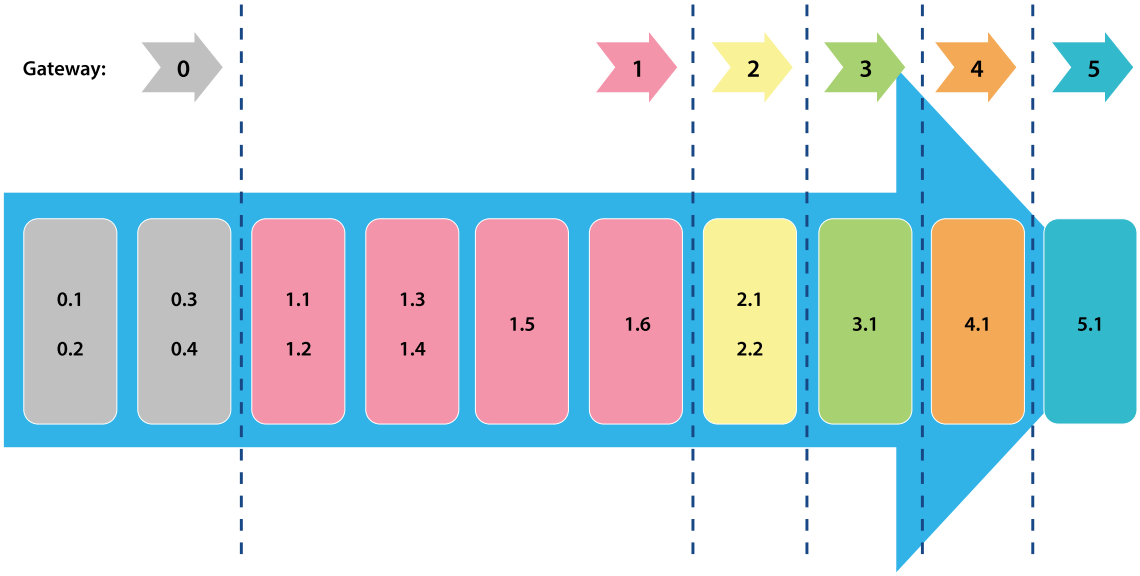
Command terms are key terms and phrases used in the syllabus content and in examination questions to indicate what is required in response to a particular command. This also suggests the type and depth of response that is expected from students.

Command terms present a particular challenge when they are translated into other languages as the subtleties of meaning can often be lost. To prevent this happening, the IB publishes its command terms in subject guides in each response language, and explains their meaning. A key responsibility when checking translated examinations is to ensure that the correct command term has been used.

Overview of examination production

Figure 26

Examination production process



Gateway 0:
Resources sign-off
0.1 Resource planning
0.2 Evaluate and select resources
0.3 Kick-off meeting
0.4 Schedule and roles agreed

Gateway 1:
Content sign-off
1.1 Draft and revise assessment
1.2 Peer review
1.3 Assessment editing meeting
1.4 External adviser ready draft prepared
1.5 External adviser report and revision
1.6 Content sign-off

Gateway 2:
Layout sign-off
2.1 Build assessment
2.2 Proofread

Gateway 3:
Usability sign-off
3.1 Scrutineer report

Gateway 4:
Translation sign-off
4.1 Translate and revise

Gateway 5:
5.1 Publish for candidates

Figure 26 describes the different stages of the examination production process. Creating an assessment is a lengthy process that takes about 18 months and involves a wide range of external experts and IB staff, including authors, subject managers, subject specialists, editors, designers, external scrutineers and linguists.

It is very important that the markscheme is developed alongside the examination. This ensures that the questions and marking are fully aligned in what they are trying to test.

Process up to content sign-off

The first stage of the process is to select who will be asked to write the examinations. While the principal examiners are usually asked, the IB also takes into account any factors that might suggest a conflict of interest, such as school connections or responsibility for running training workshops.

The roles involved in examination production are as follows.

- Author (external)
- External adviser/standardizer (external)
- Scrutineer/fluent speaker (external)
- Subject manager (IB staff)
- Production editor (IB staff)
- Copy editor (IB staff)
- Designer (IB staff)
- Translator (external)
- Language reviser (external)

Where the IB produces examinations for multiple time zones, they go through a completely independent examination production process.

Authoring and content sign-off

Authors are required to produce a first draft of the examination and markscheme. These drafts then go through several levels of review with other authors and subject managers. For some subjects, this then leads to a formal assessment editing meeting, where each examination is reviewed in terms of:

- level of cognitive demand of the questions, and overall range of difficulty of the examination
- any possible bias
- accessibility
- curriculum coverage
- confirming that the examination matches the published assessment model
- consideration of how the examination relates to past examinations and sample material
- checking that the questions are not the same as example questions in any published resources.

This meeting will produce another set of completed drafts, which will then be reviewed by the external adviser or standardizer.

The external adviser/standardizer is another subject expert who has not yet been involved in the writing of these examinations and markschemes. They provide a new perspective and are asked to comment on the level of difficulty of all the examinations taken as a whole, how they compare with previous examinations, and how well they match the curriculum. The external adviser offers suggested changes that the author and subject manager then review to inform the writing of the final drafts.

A content review is conducted internally on the final drafts with specialist checks by a copy editor and designer, in conjunction with the subject manager and author, if required.

The final content sign-off of the examination is given by the author, with the IB subject manager confirming that the examination has been reviewed properly following the IB's own quality standards.

The markscheme is not signed off at this point, but is reviewed during standardization in light of student responses.

Building and proofreading

Once the content has been signed off, the text of the questions is fixed. The designer will then build the questions in the format in which the examinations will be taken (paper-based or digital).

For traditional paper-based examinations, this means adding the front cover with the instructions (or rubric) and applying the appropriate style rules, as well as adding the various barcodes and blank pages to form the question booklet. This process also includes redrawing any diagrams to the necessary quality, or sourcing appropriate images online.

With digital assessment, the designer will need to create the examinations in the appropriate development environment. This includes the instructions/rubric and appropriate house styles.

In both cases, the completed assessments are proofread against the signed-off content by a production editor.

Usability sign-off

The scrutineer is another subject expert who has not been involved in the production of the assessment before this point. They take the assessment as if they were a student, including considering the time taken to answer the questions. The purpose of this check is to spot any errors and identify whether there is any ambiguity in the instructions or questions.

The scrutineer also reviews the markscheme once they have taken the assessment.

In some small-entry language subjects, this scrutineer role is done by a second fluent speaker of the language rather than a literature expert.

Once the formal feedback from the scrutineer has been considered, the assessment is finalized and ready to be securely sent to schools.

Quality control

It is essential that examinations are produced without errors or ambiguities as these issues can have a significant impact on the students taking them. The IB takes such issues very seriously.

There is a quality-checking process at each stage of the assessment production cycle, in addition to the formal sign-off process.

Translations

The IB offers students the chance to take (non-language) examinations in a range of response languages. Currently, most subjects are available in English, French and Spanish. Based on local agreements, other specific subjects are also offered in German, Japanese and Korean.

The need to translate examinations is considered from the earliest drafts, and the IB's authoring teams and copy editors are well aware of the issues that can emerge. The formal translation process begins upon completion of the English-language examinations.

It is critical that the process of translation does not change the meaning of the questions, in order that students are neither advantaged nor disadvantaged by taking the assessment in a particular language. The IB's external translators are subject specialists employed to ensure the technical language in the assessment remains accurate. The translation is then compared with the original examination (not just the text) by a bilingual external reviser as a final quality check.

Modified papers

The best situation is that assessments are designed so they are suitable for all students. However, there are certain requests and requirements that cannot be managed through incorporating universal design, and modified versions will need to be provided. These are managed in accordance with the *Access and inclusion policy*, and are produced and delivered to schools in the same manner as the non-modified assessments.

Examinations

- The purpose of examinations is to allow students to demonstrate knowledge, understanding and application of skills in a controlled environment that offers a consistent experience for all students and minimizes the opportunities for academic misconduct.
- In setting the examination session, the IB needs to balance the needs of all students with manageability for schools, completing the marking as quickly as possible to meet student expectations.
- The IB publishes clear rules for examination venue behaviour to minimize the opportunity for academic misconduct. These are updated in response to new technology and changing environments.
- The examination timetable is a compromise between many conflicting priorities, and globally represents the least worst option.
- Details for dealing with unexpected events, inclusive access arrangements, adverse circumstances and rescheduling can be found in the appropriate programme's *Assessment procedures*.

The period when students are taking their examinations can be very demanding and nerve-racking for them. The IB's overarching principle for this period is to minimize the stress placed on students by:

- limiting the length of the examination sessions
- where possible, avoiding clashes with other examinations the students may be taking (for example, national tests).

This must be balanced against the need for having sufficient assessment to be able to make valid conclusions on the student's performance, as well as preventing the opportunity for academic misconduct.

Preparing and managing the examination venue

Schools must conduct examinations according to a strict set of regulations laid out in the relevant programme's *Conduct of examinations booklet.*

The security of the examinations and the proper conduct of each examination is of high importance because if there is a perception among schools and stakeholders that academic misconduct has occurred, then the value of the students' results will be greatly diminished. The IB has a number of policies and procedures in place in order to maintain global confidence as set out in the "Assessment integrity" section of this publication.

Inevitably, the IB has to address incidents where a student's work is not available to the IB to mark for reasons beyond the control of the school and the student. The IB can try to mitigate the impact on the student by estimating a mark, but this is only possible if there is other evidence to draw upon. To help ensure that this is the case for paper-based examinations, the IB requires that the different answer papers (scripts) are sent to the scanning centre on different days, reducing the possibility of all the scripts being lost in transit.

Digital assessment mitigates this risk in the DPCP setting as there will be copies of work submitted for assessment available from each part of the upload process.

Developing the examination timetable

As a result of wanting to minimize the length of a session, the IB does allocate different subjects to the same time in the examination schedule. While every effort is made to consider the subject combinations students can take, the IB recognizes that a small number of clashes are inevitable, and procedures for dealing with these are described in the relevant programme's *Assessment procedures*.

Examinations are scheduled to avoid more than six-and-a-half hours of examinations in a single day, where possible. The normal pattern for examinations with multiple components relating to a particular course is to schedule the two or three examinations consecutively, starting one afternoon and finishing the next morning. Although not always possible, this arrangement is preferable to presenting all the examinations for a given course on the same day.

For most students, this spreads the examinations more evenly over the two-week (MYP) and three-week (DPCP) schedule. It allows students the opportunity to recover overnight if they feel they have not done themselves justice in a particular examination.

Principles of designing the examination timetable

The following points are a high-level summary of the principles that underlie the creation of the timetable.

It is not always possible to meet all of these principles, and in such cases a compromise needs to be achieved. The IB publishes the examination schedule at least one year before examinations will be taken.

- It is not possible to take into account public, national or school holidays (apart from 1 May) because of the number of countries and regions (territories) in which IB programmes are offered.
- Although it would be desirable not to hold examinations on either a Thursday or Friday out of respect for schools in the Middle East whose weekend falls on these days, in practical terms that is not currently possible.
- Where there are subjects with particular regional or cultural links, the IB will endeavour to take these into account. For example, Arabic literature or language examinations will not be scheduled on a Friday.
- The IB uses registration data regarding subject combinations to ensure that the minimum number of students globally are impacted by subject timetable clashes.

- Students should not be expected to be examined in two different foreign (not response) languages on the same day.
- In courses with multiple examinations, to minimize the risk of an unexpected event disadvantaging a student in all their components, they will be scheduled over at least two days.
- Students should be allowed to focus their revision for each subject in a tight window; where examinations occur on more than one day they will be scheduled on consecutive days, where possible.
- Students should have their examinations spread over the whole of the examination period, rather than over a short period. This means that, when possible, subjects with the highest candidature (for example, English, history and mathematics) will not be scheduled on consecutive days.
- For the same reason, the IB will attempt to schedule language and science examinations in each of the two weeks (MYP) and three weeks (DPCP) of the examination schedule.
- The IB's internal examination processing requirements mean that certain subjects will normally be scheduled early in the timetable, in particular, large-entry subjects.

Dealing with unanticipated events

The relevant programme's *Assessment procedures* sets out how to deal with most unanticipated events. If in doubt, contact IB Answers, who will be able to provide advice.

Inclusive access arrangements and adverse circumstances

These topics are dealt with in detail in the "Prioritizing fairness for all" section of this publication.

The IB believes that all students should be allowed to demonstrate their ability under assessment conditions that are as fair as possible. Standard assessment conditions may put students with learning support requirements at a disadvantage by preventing them from demonstrating their level of attainment. Inclusive access arrangements may be authorized in these circumstances, which may be special arrangements in terms of taking the examination, or modifications made to examinations.

Events that occur during the examination session that are beyond the control of the student, and are likely to have a negative impact on their performance, are known as adverse circumstances. These circumstances include temporary illness or injury, severe stress, exceptionally difficult family circumstances, bereavement, or events that may threaten the health or safety of a student. Adverse circumstances may also include an event that affects the whole school community, such as civil unrest or a natural disaster. They do not include shortcomings on the part of the school.

For details about these arrangements, please refer to the relevant programme's *Assessment procedures*.

Rescheduling of examinations

Rescheduling an examination poses a major risk to the integrity of the examination process. It means students will be sitting the examination before or after other students take it, and could easily lead to academic misconduct. Examinations will never be allowed to be rescheduled a day before they were due to be taken as the risk of a breach that would affect the majority of students taking the examination is too great.

The only circumstances under which the IB will authorize a student to take one or more examinations at a different time and/or on a different date from the IB examination schedule, and the processes for requesting rescheduling of an examination, are defined in detail in the relevant programme's *Assessment procedures* and *Preparation for examinations policy* available on the Programme Resource Centre (PRC).

Marking

- Successful marking means that students are given a consistent and accurate score that reflects the quality of their work.
- The considered views of the principal examiner are correct, and every other examiner must reproduce this in their marking.
- Marks and grades are not the same thing; students may be able to get more marks on an easier examination but they should still receive the same grade.

Marks and grades

An important aspect of carrying out, and using, summative assessments of students is to understand the difference between marking their work and grading their work.

- In marking, a student is given credit for the work they have produced against a markscheme or similar framework. This is an indication of the degree of the assessment task they got right. The mark itself has no other meaning.
- In deciding a grade, the examiner is making a judgement on the quality of the student's work against a defined standard, which will take into account the difficulty of the task as well as the proportion of the task that was completed. The grade, therefore, has some meaning or relevance and is usually intended to be comparable with performances on other assessments.

It might be possible to demonstrate a high grade from getting only a small proportion of a very difficult question correct; equally, it may be impossible to demonstrate the same grade by correctly answering many trivial questions.

As discussed in later sections, it is not necessary for the standard described by the grade to be explained by reference to what the student has attained, although this is the approach taken by the IB. There are other perfectly consistent and well-respected systems where the standard is based on how the student performs relative to their peers.

In its assessments, the IB generally uses marks as an indication of overall performance. It then looks at how students with this number of marks performed to determine a boundary point (grade boundary) at which students with more than that number of marks are awarded a particular grade. This process is explained in more detail in the "Grade awarding and aggregation" section.

Approaches to marking

As the IB uses different assessment instruments, it is essential that the right type of marking approach is selected for each assessment. The decision about the marking approach is taken during the curriculum development process.

Analytic markschemes

Analytic markschemes are used for assessments where there is a narrow range of expected answers from the students.

These markschemes can give specific instructions to examiners about how to break down the total marks available for a question between different parts of the answer. Even with structured questions where highly specific answers are expected, markschemes must provide examiners with sufficient information for them to mark consistently, considering the different kinds of approach that students might adopt and the common errors that they might make. It is inevitable that examiners will need to use their professional

judgement in allocating marks to unexpected responses or alternative valid answers, but markschemes must provide as much guidance as possible in how to exercise that judgement.

Students do not always follow predictable patterns in what parts of a question they get right or wrong. This is an issue particularly with extended structured questions, where a mistake in an early part could have an impact later on. The IB designs such questions so that if a student makes a mistake in the early part of the question, the rest of that question does not become inaccessible to the student. Nevertheless, even analytic markschemes need to provide explicit guidance on how to mark particular kinds of incorrect answer, and how to deal with following through students' working when they have made a mistake in part of a question.

Assessment criteria

For other assessments, where a detailed markscheme wouldn't be appropriate, assessment criteria are applied instead.

Assessment criteria concentrate more on the type of performance that students are expected to demonstrate, regardless of the specific details of the response. The different levels of expertise that could be demonstrated are reflected by level descriptors.

The IB uses a "best-fit" model in the application of criterion level descriptors. The examiner applying an assessment criterion must choose the achievement level that overall best matches the piece of work being marked. It is not necessary for every detailed aspect of an achievement level to be satisfied for that level to be awarded, and it is worth noting that the highest level of any given criterion does not represent perfection.

In most components that are marked using assessment criteria, multiple criteria are used rather than a single criterion. It is important that the criteria are independent of each other. It would not be fair for a student to gain credit in several different places for the same aspect of their answer.

Usually, the same assessment criteria are used every year, even though the questions asked of the students are different. This is because the underlying nature of what is being assessed remains unchanged. The different question generally relates to the specific details that are not explicitly set out in the criteria. The principal examiner will often provide marking notes that offer guidance to examiners on how the assessment criteria should be applied to each question, as well as examples of relevant details.

Holistic criteria: Markbands

Sometimes, it is not appropriate to separate out the different assessment criteria to mark a piece of work. This usually happens when it is impossible to have distinct criteria that are independent of each other. In such cases, markbands are used instead of separate criteria. The markbands, in effect, represent a single holistic criterion applied to the piece of work, which is judged as a whole. Because of the requirement for a reasonable mark range along which to differentiate student performance, each markband level descriptor will correspond to a range of marks.

Table 7

An example of an IB markband

Marks	Level descriptor
0	The work does not reach a standard described by the descriptors below.
1–3	• Little understanding of the issue is demonstrated. Subject-specific vocabulary is not used or is consistently used inappropriately. • The explanation of the issue is minimal. Points made are superficial and frequently unclear. • The response is descriptive. Any analysis present is superficial or incoherent. There is little or no discussion of different points of view. Where a conclusion is included, this is very superficial or is not consistent with the rest of the response.
4–6	• A basic understanding of the issue is demonstrated. Subject-specific vocabulary is used, but often inappropriately.

Marks	Level descriptor
	• The explanation of the issue is basic and underdeveloped. Points are often imprecise or vague, and it is often unclear what the response is trying to convey. • There is limited analysis present, and overall the response is more descriptive than analytical. There is limited discussion of different points of view. A simplistic conclusion is included.
7–9	• Some understanding of the issue is demonstrated. Subject-specific vocabulary is used, sometimes appropriately. • There is a satisfactory explanation of the issue, although the explanation lacks clarity and development in places. Relevant points are made but lack accuracy and detail. • The response contains analysis, although this analysis lacks development. There is some discussion of different points of view. A conclusion is included.
10–12	• Good understanding of the issue is demonstrated. Subject-specific vocabulary is used, mostly appropriately. • The explanation of the issue is clear but is in need of further development. Points made are relevant and accurate but lack detail. • The response contains critical analysis, although this analysis lacks development. There is discussion of different points of view. The response argues to a conclusion that is consistent with the arguments presented.
13–15	• Very good understanding of the issue is demonstrated. There is accurate and precise use of subject-specific vocabulary. • The explanation of the issue is clear and well developed. Points are relevant, accurate and detailed. • The response contains well-developed critical analysis. There is critical discussion of different points of view. The response argues to a reasoned and clearly stated conclusion that is consistent with the arguments presented.

The descriptors themselves tend to be quite lengthy, covering a range of potential qualities evident in students' work, and will relate directly back to the course objectives. As with assessment criteria, a best-fit approach is used. Examiners need to make a judgement about which mark to award from the possible range for each level descriptor, according to how well the student's work fits that descriptor. For example, one markband level may cover 7–9 marks. The examiner will award a mark from that range according to how well the student's work fits the relevant level descriptor from the markband scale.

Automarking

A small number of components that consist of multiple-choice questions are automarked. This is the process of using technology to assess students' work against a pre-defined markscheme. For these multiple-choice questions, the markscheme states whether an answer is objectively right or wrong. Markschemes for these components are extremely clear and there is no need for any examiner judgement to be applied.

Marking that requires examiner judgement

While some assessments can be automarked, the majority of IB assessments need to be marked by examiners, as analytic markschemes or marking criteria need to be applied.

Examiners mark students' work electronically using a piece of software known as a "marking tool". This allows them to put ticks, marks and notes on the electronic copy of a student's work, including audio and visual material. This opens up a much broader range of work that can be submitted as evidence, for example, recordings of performances or oral examinations.

For more information on how the IB marks, please watch the video "The two methods of marking: How marking is carried out by examiners and teachers".

Standardization

The process of standardization covers several elements of the assessment procedure. The purpose of standardization is to prepare examiners to mark with reliability.

For outcomes to be reliable, it is important that each script receives the same mark based on the quality of work it contains, irrespective of which examiner marks it. To this end, examiners must demonstrate an understanding of the marking standard set by the principal examiner for a component, and align with it, before they are able to mark in a live IB examination session.

The first step in the process of standardization is to recruit examiners with the appropriate skills and experience to mark. The examiners are then given training to understand marking principles and practices.

Once an examination has been taken, the senior examining team, led by the principal examiner, meet to discuss the markscheme and to review sample scripts. The aim of this standardization meeting is to ensure that the markscheme adequately covers the anticipated range of answers that examiners are likely to encounter. A markscheme is unlikely to cover every possible response that a student may provide, but reviewing sample scripts at the meeting enables the principal examiner and senior examining team to produce clear instructions and guidance to inform examiners of the type of response required for each mark, based on the work seen so far.

The outcome of the standardization meeting is a set of scripts marked definitively by the principal examiner. These scripts are used to train other examiners to mark to the same standard as the principal examiner. Only examiners who have successfully demonstrated alignment with the marking standard of the principal examiner are permitted to mark scripts in a live examination session, and their marking standard is repeatedly checked throughout the marking period with robust quality assurance procedures.

The quality model

It is important that whichever examiner marks a student's work, the student should receive the same outcome as if the principal examiner had marked their work.

After the assessment has been taken and some student responses are received, the principal examiner holds a standardization meeting with the other senior examiners. The senior examiners agree on the marks for a number of responses that will be used to set up the quality model. There are three categories of definitively marked scripts: practice, qualification and seed scripts. Together, they comprise the quality model.

Practice

The purpose of practice scripts is to support examiners in learning the marking standard that has been set by the principal examiner. They should show how to mark typical examples of scripts and identify any common situations where a specific understanding of the markscheme may be required.

A good set of practice scripts would include a range of student performances, from low marks to high, to help examiners recognize where they might award marks and what a good (or bad) answer looks like.

After reviewing the markscheme or marking criteria, and any marking notes that the principal examiner has provided after the standardization meeting, examiners can attempt the practice stage. The practice scripts demonstrate the principal examiner's marking standard. While examiners are marking the practice scripts, they can ask their team leader for guidance if needed.

Once examiners have completed the practice scripts, they should be confident that they understand and can apply the principal examiner's marking standard, and they can then start the qualification process.

Qualification

The purpose of qualification scripts is to provide examiners with the opportunity to demonstrate that they can mark to the correct standard. Following successful qualification, examiners will then be able to mark in a live examination session.

There are usually five qualification scripts in a set. If an examiner does not award the same marks as the principal examiner while attempting qualification, they are given feedback on what mark should have been awarded, and why. Examiners are then offered a second opportunity to demonstrate that they now understand the marking standard using a different set of qualification scripts.

A good set of qualification scripts would include a range of student performances, from low marks to high, and examples of scripts that require the examiner to have understood any specific instructions in the markscheme that will assist them in dealing with anticipated challenges that the principal examiner and senior examining team feel they may encounter while marking. These will be based on the findings from the review of scripts during the standardization meeting. Qualification scripts are not selected to try to catch out examiners; they represent the types of response that examiners are likely to see when they mark.

In this stage, there are two opportunities for examiners to demonstrate they have understood the principal examiner's marking standard and can apply it to work submitted for assessment. Examiners have to successfully complete a qualification set.

Each qualification set is made up of a number of definitively marked scripts. Examiners are not allowed to discuss the scripts with their team leader while they are completing a qualification set. If an examiner's marks in the first qualification set are within tolerance, they progress to live marking.

However, if the examiner's marking is outside of tolerance, they are given feedback from their team leader about how to align their marking with the principal examiner's standard. After they have reflected on this, they have a second opportunity to show they understand the marking standard.

If an examiner's marking in the second qualification set is within tolerance, they are then allowed to start live marking.

Live marking and seeds

Seed scripts are a part of the quality assurance process. They are used to monitor examiners' performance and demonstrate that they are continuing to mark to the correct standard.

Over time, an examiner's alignment with the standard set by the principal examiner can start to drift. For this reason, the IB periodically checks their marking using seed scripts, which have already been marked definitively by the principal examiner. Seed scripts look the same as other scripts, so an examiner is unaware that they are marking a script for quality assurance reasons. In general, an examiner will be asked to mark one seed script at random in every 10 scripts they mark.

If the examiner gives the same mark as the principal examiner for a seed script, they can continue with their marking. However, if they demonstrate they are no longer marking to the required standard by awarding an incorrect mark, there is an intervention by the IB.

Initially, feedback and guidance are provided by senior examiners to allow the examiner an opportunity to re-establish their ability to mark to the standard. If they are unable to do so despite the extra support, they are prevented from marking any further scripts.

Marking tolerances are established for each component, which permits a certain degree of variation from the definitive standard. For more information, refer to the "Tolerances" section.

Indicators of successful standardization

A successful standardization exercise results in:

- engaged senior examining teams that are able to assist examiners with marking queries and ensure marking consistency
- clear, unambiguous instructions, in the form of the markscheme or marking notes, that clearly explain the rationale for marks awarded

- definitively marked practice, qualification and seed scripts for training and quality assurance purposes.

Tolerances

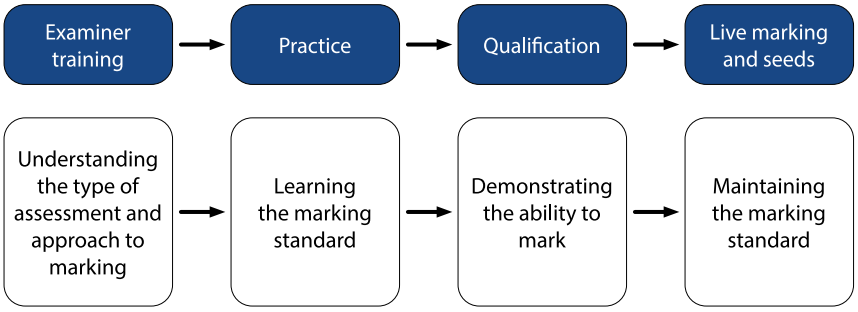
The IB believes in asking questions that test what is important, not just what is easy to mark. For certain types of questions, where the answer is either right or wrong, it is reasonable to expect an examiner to give exactly the same mark as the principal examiner. For other types of questions, especially those that test understanding or analysis, a degree of judgement is involved. In practice, this may mean that two examiners with the same expertise and understanding of how a question should be answered successfully will not award exactly the same marks as each other for the same piece of work. They may differ by 1–2 marks.

The IB uses the idea of tolerances to reflect these legitimate differences in opinion when reviewing examiner performance. For example, if the principal examiner gave an essay a mark of 46 and the question had a tolerance of 2 marks, then the IB believes that any examiners who give it a mark between 44 and 48 have shown they are marking to the appropriate standard.

The IB defines how close examiners need to be to the definitive mark (the principal examiner's mark) when they mark these scripts. This allowable difference is the tolerance. When reviewing questions with several parts or criteria, the IB monitors both the difference in overall mark and also differences for each part or criterion. An examiner must be within tolerance for both aspects to show they are marking to the correct standard.

Figure 27

Aligning marking



Examiner training

Before an assessment session starts, training is available to examiners. The purpose of the training is to help examiners understand the type of assessment and the general approach to marking it. The training does not contain specifics about examinations for future sessions.

Challenging and unusual scripts

If examiners come across scripts they cannot decide how to mark, they contact their team leader to ask for advice. If they still do not believe they can mark it fairly, they can send it to an examiner more senior than themselves for marking. Particularly problematic scripts will be brought to the attention of the principal examiner, who can provide the definitive decision on how to mark the script.

In a similar way, any unusual scripts, or the scripts of students who have answered modified question papers, will be marked by the principal examiner. It is the job of the principal examiner to carefully balance the marking to ensure it is equivalent to the same standard that is applied to other students.

School connections

The IB's principle is that examiners should not mark their own students' work.

To manage this, the IB instructs all examiners to disclose any connections they have to students and schools so that they do not receive students' work from schools with which a conflict of interest exists. For further information on this process, please refer to the "Conflicts of interest" section of this publication.

Examiners' notes and annotations

The purpose of IB summative assessment is to measure a student's performance. Examiners are asked to focus solely on marking students' work to the required standard. The IB only asks examiners to make comments when it helps them in carrying out their marking.

Writing formative feedback for either students or teachers requires the examiner to determine the correct mark for a piece of work and then try to explain how the work could have been improved. The IB acknowledges that this second task is at the heart of what good teaching means, and it is not a trivial task. It requires time and thought, which will draw the examiner away from their core task of marking to a consistent standard. In simple terms, the IB wants examiners to do one task—marking—to a high standard, rather than two tasks—marking and feedback—to a lower standard.

An examiner can also only make their judgement on the one piece of work they have available, and experienced teachers will draw upon a wide range of information when deciding how to offer feedback to a student. Therefore, the quality of any examiner feedback will suffer from having less insight than that of the teacher.

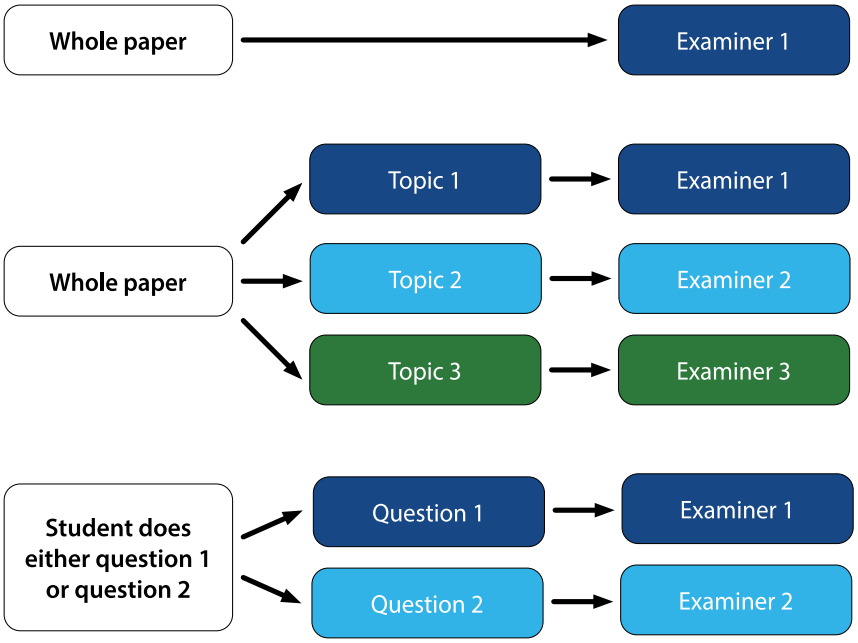
For all these reasons, the IB is very clear to its examiners that they should only mark the student's work according to the correct standard, and that they should not add comments to provide feedback to the student or teacher.

To assist with marking processes, examiners are required to indicate clearly where marks have been awarded and, if there could be ambiguity, to clarify with appropriate comments. This supports the IB in checking standards and also provides transparency for schools on where marks have been awarded.

Question item groups

Figure 28

Dividing papers into question item groups



The idea behind question item groups is that it is easier for examiners to mark the same question many times, rather than marking all the questions on a student's whole script and then starting again for the next student.

Each question item group will have its own practice, qualification and seed scripts, and examiners will need to prove they can mark to the required standard on each question item group. While this may seem challenging, it means that an examiner who cannot grasp the standard on one particular question of an examination can still mark all the other questions, rather than being stopped because they cannot mark the whole examination to the necessary standard.

Aggregation

- Aggregation is the process of combining components to generate an overall result.
- In order for each component to contribute in the correct proportion to the final mark (weighting), it may be necessary to scale the component marks.
- The IB uses a "compensation model", where students can offset poor performance in one question or component with high attainment in another.
- A student's final subject grade is determined from the aggregation of component marks, and not from component grades.

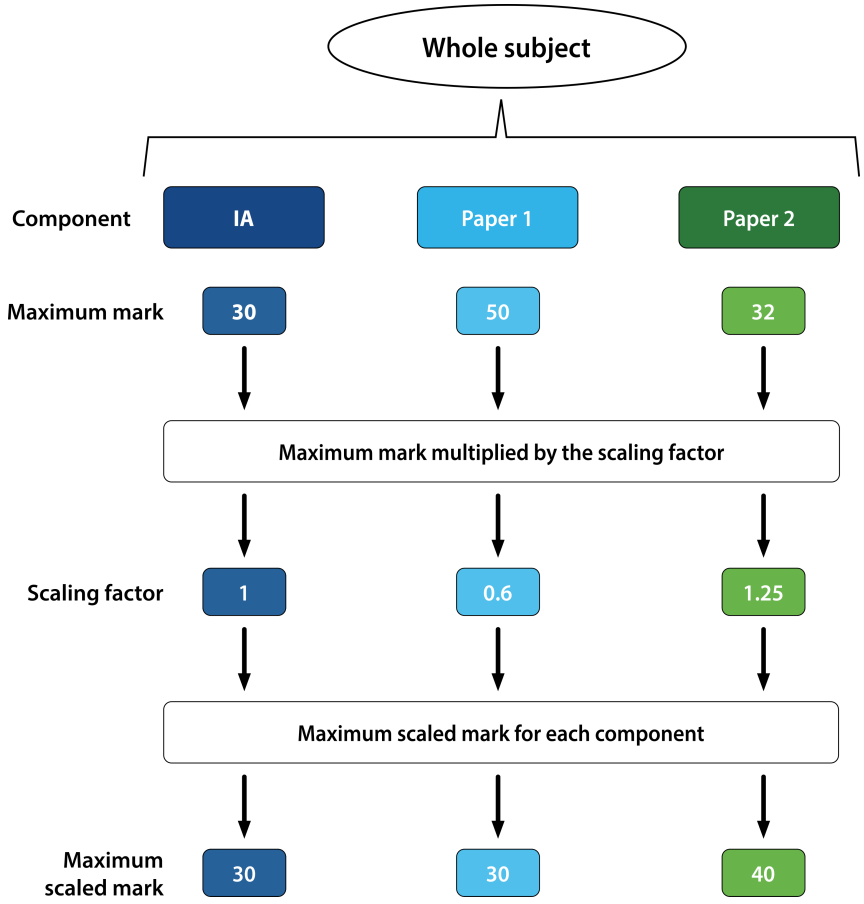
Aggregation is the process of combining marks (and boundaries) from the different components to form a final mark or overall grade boundary. To achieve this, overall component marks (or boundaries) may need to be scaled.

Scaling is carried out to preserve the desired weighting for each component in terms of its contribution to the overall assessment for the course. It means multiplying or dividing component marks so that they contribute correctly to the overall total for the subject. The same applies to the grade boundaries set for the component, which would have to be determined initially out of the maximum marks for the components.

The maximum scaled mark for each component is the same as the weighting given to the component in the subject guide.

Figure 29

Example of scaling in a subject with three components



The concept of weighting is to reflect the relative importance that the IB places on the elements being assessed in contributing to the final outcome. For example, if a component primarily tests interpretation of data or sources, and has a weighting of 30%, this implies that, compared with the other objectives of the course, interpretation represents about 30% of what is important. Often, several components will test similar objectives and so this calculation is less meaningful.

A secondary, but nonetheless important, aspect of weighting is that it allows the IB to set the total number of marks in an assessment that is appropriate to the tasks and the marking criteria, rather than trying to force it artificially into an overall total. Those totals can then be aggregated to arrive at a final grade.

Aggregation in assessment involves combining marks from different components to determine a student's final grade. This approach allows for flexibility in setting marks based on task difficulty and criteria and the intentions of the assessment design, rather than adhering to a fixed total indicated by component weightings. It means that while component weighting is a consideration, the actual outcomes may be different to the intended outcomes due to, for example, unforeseen comparative ease or difficulty of an assessment or item. The IB's approach does not include aligning individual marks to match predetermined weightings exactly; the overall outcomes are what matter most. This method, known as the "compensation model", enables students to offset poor performance in one area with strong performance in another, focusing on the total mark rather than individual component marks. The final grade is, therefore, derived from the sum of all component marks, and not from averaging component grades, which means students with the same component grades might receive different final grades.

Moderation

- Moderation is the process of checking teachers' marking standards, not the process of re-marking students' work.
- Successful moderation means that students would receive the same IA mark even if they had gone to another school on the other side of the world. The IB refers to this as the "global standard".
- The evidence used in moderation is not just the quality of the students' work, but also the teachers' explanation of why they have awarded marks.

Moderation, using dynamic sampling, is used for IA in the DP and CP, and for ePortfolios in the MYP.

Defining moderation

In many cases, the trait the IB wishes to assess in a student cannot be tested in a formal, time-limited examination. In such cases, the most valid approach is to ask the teacher to carry out the testing with an IA. The section "The role of classroom-based assessment and internal assessment" sets out why this is an appropriate approach.

While this produces meaningful results, it also creates a risk that different teachers have different interpretations of what the marking standard is. Two teachers in two different schools might award the same piece of work two different marks. The IB has a comprehensive process for training and testing examiners so that they all have a common understanding of the required standard, but this is not feasible for all teachers.

Instead, the IB has confidence that its teachers are marking to a consistent standard for all their students, so it only needs to apply an adjustment to their marks, if necessary, to bring them in line with the global standard. The IB does this by asking for a sample of teachers' marking and comparing this to the principal examiner's standard. From the data this comparison provides, where necessary, the IB produces a mathematical formula that adjusts the marks provided by each teacher. Based on a statistical comparison between the two sets of marks, an adjustment is made to the teacher's marks for all students at the school (for that component), if required. If the teacher is consistently under- or over-marking, this adjustment will be the same for each of the teacher's marks, but if the teacher is under- or over-marking either at the top or bottom of the mark range, this adjustment may vary across the range of the teacher's marks.

Important points to remember

- The aim of moderation is to check how accurately and consistently a teacher has applied the assessment criteria in their marking of the students' work.
- As a result of moderation, a school's marks may be lowered or raised. They may also remain the same.
- A moderation factor does not mean that the teacher's marking is of poor quality, only that it is not consistent with the global standard.

For practical reasons, the IB moderates schools rather than individual teachers, and so it is very important that all teachers in a school ensure they are marking to the same standard. The moderation factor is applied to all teacher marks, so the sample must be a fair representation of the school's marking to ensure that any adjustment is appropriate.

Considerations for a coordinator when reviewing IA marks that will be moderated include the following.

- Are all subject teachers (who are responsible for providing IA marks) in the school marking to the same standard?
- Are teachers consistent in marking all students to the same standard?

- Are teachers providing clear explanations for why they awarded a particular mark?

Further guidance for teachers on applying assessment criteria accurately can be found in the "Feedback to schools" section of this publication, and in the current MYP, DP and CP *Assessment procedures* publications.

For further details of the technical aspects of the calculation of moderation factors, see "Moderation of internal assessment: A closer look" in the appendix to this publication

Moderation and expectations of teachers

Work submitted for assessment is internally assessed by teachers and externally moderated by the IB.

Teachers should familiarize themselves with the relevant subject guide, which will provide details of the IA criteria. Each assessment criterion has level descriptors describing specific achievement levels, together with an appropriate range of marks. The level descriptors concentrate on positive achievement, although for the lower levels failure to achieve may be included in the description.

Teachers must judge the internally assessed work against the criteria using the level descriptors and aided by the clarifications. The criteria must be applied using a best-fit approach. When a piece of work matches different aspects of a criterion at different levels, the mark awarded should be one that most fairly reflects the balance of achievement against the criterion. It is not necessary for every single aspect of a level descriptor to be met for that mark to be awarded. The highest level descriptors do not imply that the student's performance is faultless.

Where there are two or more marks available within a level, teachers should award the upper mark if the student's work largely satisfies the qualities described (the work may be close to achieving marks in the level above). Teachers should award the lower mark if the student's work demonstrates the qualities described to a lesser extent (the work may be close to achieving marks in the level below).

Only whole numbers must be recorded. Partial marks (fractions and decimals) are not acceptable.

The criteria should be considered independently. A student who attains a high achievement level in relation to one criterion will not necessarily attain high achievement levels in relation to the other criteria. Similarly, a student who attains a low achievement level for one criterion will not necessarily attain low achievement levels for the other criteria. Teachers should not assume that the overall assessment of the students will produce any particular distribution of marks.

In some subjects, level descriptors may refer to specific command terms. These command terms are to be interpreted as indicated in the relevant section of the guide.

Teachers should be aware that when a moderation factor is applied to their internally assessed work it does not mean that they followed the marking procedure incorrectly. The adjustment is simply to bring their mark in line with the agreed global standard. The subject report for each subject, issued after each examination session, will often have useful information for teachers about IA performance. Teacher support materials are also available on the PRC.

The same moderation factor is applied to all work submitted for assessment from each school. Accordingly, if different teachers share responsibility for IA in the same subject within a school, it is essential that internal standardization takes place so that the marking standard within the school is as aligned as possible. This may include discussion of the assessment criteria and a review of work submitted for assessment before any marking takes place.

Teachers and students must understand concepts that relate to academic integrity, especially authenticity and intellectual property. Students' work must be entirely their own and prepared according to the requirements listed in each subject guide. Where collaboration is permitted, the differences between collaboration and collusion must be made clear to all students. For further information, refer to the section "Nurturing an ethical mindset".

Teachers wishing to access a summary of this information can print the poster "Marking internal assessment: Expectations of teachers in the moderation process" from the "Printable resources" appendix to this publication.

Selection of student work

- All students should be given a fair mark. Therefore, any student's work is appropriate to be used as an example of the teacher's standard for moderation.
- Teachers may have different ideas of what constitutes very good or very poor work, so it is important that a school's sample covers the full range of marks awarded.
- To ensure transparency and fairness between all schools, the IB must be able to see evidence of all work that is contributing to a student's final mark.

To ensure transparency and remove any perception of academic misconduct, it is important that the IB, rather than the school, identifies the work that is to be sampled for moderation. The IB uses broad guidelines to ensure that the moderation factor is as reliable as possible and, within the limitations of these guidelines, that the actual selection of students is random.

The first principle is that the IB should use the smallest possible sample to obtain a reliable moderation factor for any school. This minimizes the burden on the school and also the costs to the IB, which would then be passed on to schools through examination fees. It is, however, reasonable for the number of pieces of students' work in a sample to be different for different schools if the IB requires more examples of teacher marking to determine a robust moderation factor.

The second principle is that the IB must be confident that its moderation factor is fair to all students across the mark range. From experience, the IB knows that teachers can have different expectations at different points of the mark range, and that a teacher who is more generous than the global standard for poor-quality work may be harsher than the global standard for high-quality work. For this reason, the IA sample is carefully selected to ensure that the entire mark range of the school is appropriately represented.

The IB tends not to select students for the moderation sample who have attained full marks. This allows students in the higher mark range the possibility to be moderated upwards if a teacher is too harsh in their marking. Also, work from a student who had been awarded 0 marks would not usually be selected.

Irregular situations

Moderation is designed to check that teachers are marking to a global standard, even if a piece of work is particularly challenging to mark. In such situations, the teacher may well need to give a more detailed justification of why they have awarded that mark. For example, if a student has received additional assistance from the teacher to complete their work, the teacher would explain this in their marking comments, and the examiner would take this into account when reviewing the teacher's marking of the student's work.

Failing to find a moderation factor

In some cases, it may not be possible to calculate a moderation factor using the submitted sample work. This happens if the difference between the examiner's marks and the teacher's marks are inconsistent, or the examiner's marks indicate the teacher's marking is too generous or too harsh compared with the global standard. In these cases, the IB will request further work from the school so that it can be sure a fair moderation factor can be applied. This is why all students' work must be available until the close of the examination session. In rare cases, if a fair moderation factor cannot be applied, examiner marks will be awarded.

Dynamic sampling

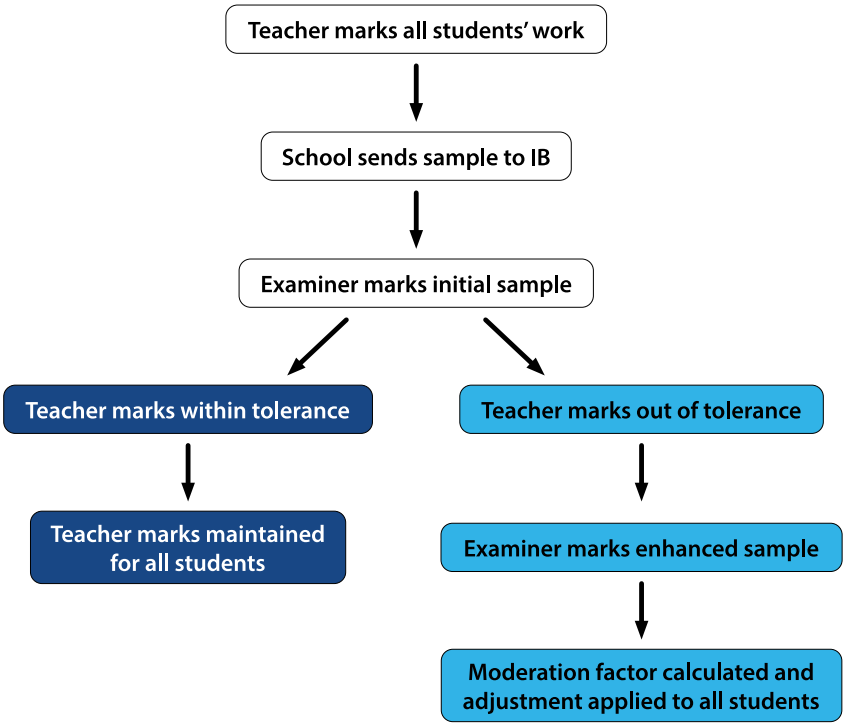
- Moderation using dynamic sampling means that if a teacher is marking within tolerance, the IB accepts their marks.
- If any of the teacher's marks are outside of tolerance, the IB applies a moderation factor.

- Every examiner who is checking teachers' IA marking is monitored for quality assurance.

The principle of moderation using dynamic sampling means that if a teacher demonstrates they understand the global standard by marking within tolerance on their sample, it is accepted that all their marks are to the required standard. If they are outside of tolerance, then a moderation factor is calculated and applied. This is explained in figure 30.

Figure 30

Moderation using dynamic sampling



For moderation to be fair, examiners need to understand the global standard set by the principal examiner and to moderate to it. Examiners undertaking moderation receive training in the global standard and are then monitored to ensure they are moderating to this standard. As described in the "Standardization" section, the principal examiner prepares three types of definitively marked IA work.

- **Practice scripts**, to explain the global standard
- **Qualification scripts**, to check examiners understand the global standard
- **Seed scripts**, to check examiners are maintaining the global standard

The same examiner will review all the IA work from a single school. However, examiners are presented with work from a range of schools in no particular order. This mitigates against examiners forming an opinion on whether a teacher has over- or under-marked from the work of the first student they see, and then looking for this pattern in the rest of a school's sample. It also means seed scripts can be included without it being obvious that they are seeds. Once it has been determined whether a moderation factor is required, examiners can be asked to provide summary comments on a school's marking.

The quality model in dynamic sampling is more complex than with other marking. The IB also needs to check that examiners are confident reviewing teachers' marks that may be too harsh, too generous or correct. Figure 31 shows the range of quality model IA scripts required to cover all scenarios.

Figure 31

Range of quality model IA scripts required

	Scenario	
Examiner mark near top of mark range	1	Teacher mark too harsh
	2	Teacher mark at global standard
	3	Teacher mark too generous
Examiner mark in middle of mark range	4	Teacher mark too harsh
	5	Teacher mark at global standard
	6	Teacher mark too generous
Examiner mark near bottom of mark range	7	Teacher mark too harsh
	8	Teacher mark at global standard
	9	Teacher mark too generous

For more information on how the IB moderates, see "The two methods of marking: How marking is carried out by examiners and teachers" (video).

Grade awarding and aggregation

- Grades should mean the same, whichever session a student takes their examination in.
- The grade award process decides how to convert between marks and grades to ensure that this is the case.
- Grade boundaries are determined using a range of evidence, including both cohort results and expert judgement of work submitted for assessment.
- It is the overall grade that a student receives that is the most important aspect, not any individual components.

It is important to remember that marks and grades are not the same thing. For more details on why this is the case, see the section "The difference between marks and grades". The grade award process is how the decision is reached on how marks correspond with grades in a given instance.

This decision is made over several days of discussion between the principal examiner, chief examiner, senior examiners and the IB. It draws upon a range of evidence (described in more detail in the section "Evidence used in the grade award process") to reach a conclusion that is fair to students this year, but also to students who have taken the subject in previous years. Finally, recommendations are made by the chief examiner to the IB on what the outcomes for this examination session should be.

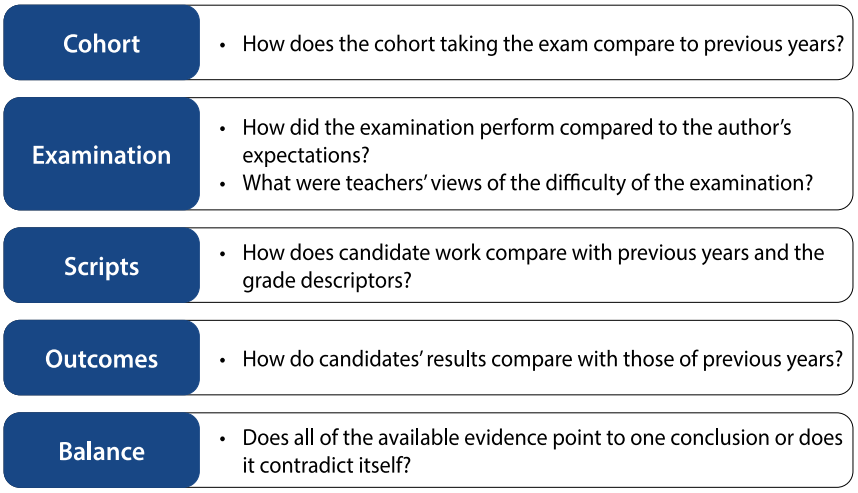
The main outcomes of the discussion are the grade boundaries, the minimum number of marks that are required by a student to obtain each grade. As it is not practical to make a detailed judgement for every single grade in every subject, the IB asks its examiners to make a recommendation on several "judgemental grade boundaries" and calculates the remaining boundaries arithmetically. For further details, please refer to the section "Judgemental and interpolated grade boundaries".

Grade boundaries will not necessarily be the same for each examination session. This is because the questions that students are asked to complete will be different for each iteration of an examination, and grade boundaries need to vary to reflect easier or more difficult tasks. The IB makes every reasonable effort to ensure that its examinations are of the same level of difficulty every year. However, because of their high-stakes nature, it does not trial examinations before they are taken because of the risk of questions being made public.

In making the decision about where to place the grade boundaries, examiners and the IB consider the aspects shown in figure 32.

Figure 32

Aspects considered during the grade award process



In many circumstances, the IB has several assessment components that are combined to give an overall grade. Further detail on this is given in the "Aggregation" section of this publication. The most important aspect that examiners must keep in mind during the grade award is that it is the **overall result** that is most important. If necessary, individual component grade boundaries can be less perfect in order to achieve a fair overall outcome.

Formally, the purpose of the grade award process is:

- to establish the point(s) in the distribution of work submitted for assessment where there is a change in which grade descriptor best describes its quality
- to determine the marks for each component that are the grade boundaries
- to ensure that the combination of these grade boundaries at the overall subject level represents a fair awarding of grades.

A grade award process has been successful if:

- script judgement and outcome evidence are in broad agreement, taking cohort information into account
- any significant variation in school performance can be explained
- the chief examiner and IB chief assessment officer are confident that assessment standards have been maintained.

Judgemental and interpolated grade boundaries

The **judgemental grade boundaries** are those that are recommended by the chief examiner based on discussions during the grade award process. For the MYP, DP and CP, these judgemental grade boundaries are the 2/3 boundary, the 3/4 boundary and the 6/7 boundary. The remaining boundaries —1/2, 4/5 and 5/6—are calculated arithmetically based on the judgemental grade boundaries and are known as the **interpolated grade boundaries**.

If there was a significant shift in the proportion of students receiving a grade 5 or a grade 1, this would need to be discussed in the context of the cohort. This may lead to either a reconsideration of the judgemental grade boundaries or, in exceptional circumstances, to a review of work submitted for assessment at these boundaries.

Evidence used in the grade award process

Grade awarding is an evidence-based process that draws upon a range of information including:

- teacher feedback on the assessment
- expert judgement from examiners on examples of students' work
- a review of statistical information, comparing this year's student outcomes with those from previous years.

No one type of evidence is more important than any other: all must be balanced equally in coming to a conclusion.

Considering this year's cohort

The first task of the grade award process is to consider how similar the cohort taking the assessment is to that of previous years. If the schools with students taking the assessment this year are largely the same as in previous years, then any differences in performance are likely to be due to the difficulty of the examinations. If this is not the case (for example, if there is a large number of new schools taking the subject for the first time, or a high proportion of students are retaking the assessment), then the IB might decide any differences in performance need to be reflected in grade outcomes.

Examples of the factors that will be considered when comparing the cohorts taking the assessment include:

- changes in the number of students taking the assessment
- changes in the proportion of students taking the assessment in English, French and Spanish, and other languages where appropriate
- the number of new schools and the number of students in those new schools
- any changes in the options taken by students.

Feedback on the assessment

The next task is to consider how the assessment performed relative to expectations. If a particular question proved far more challenging than expected, or questions that were expected to discriminate between grade 6 and grade 7 students failed to do so, then the grade awarding team need to take this into account when determining grade boundaries.

They base this discussion on teacher feedback about the examinations and statistics on performance on individual questions or items (see figure 33). The team also have summarized reports from individual examiners and are able to draw upon their own experiences of marking students' work during the session.

Figure 33

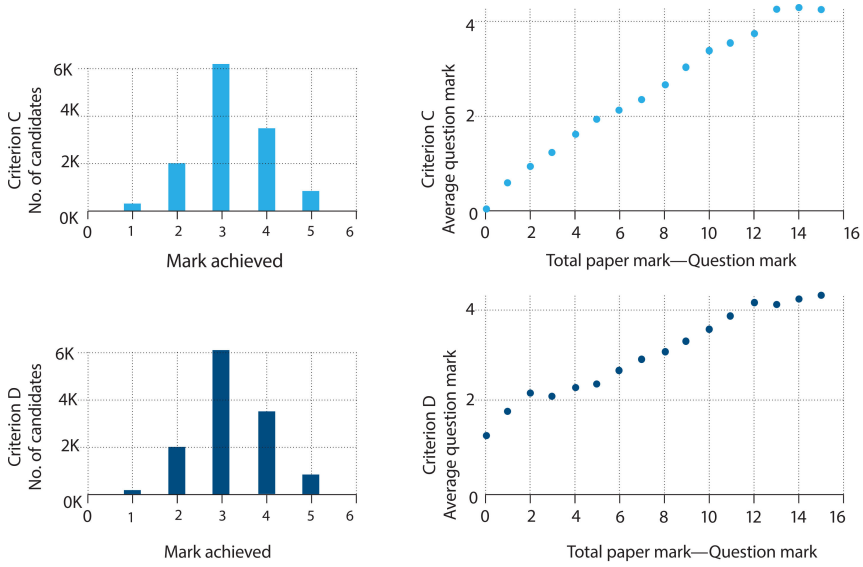
Example of item-level statistics available to examiners

Item summary statistics

Question	Question attempted	Total candidates for exam	Proportion of candidates attempted	Average question item mark	Highest mark achieved for item	Population standard development of question item mark
01 Criterion A	12,071	42,042	28.71%	2.7	5	0.95
01 Criterion B	12,071	42,042	28.71%	2.6	5	0.94
01 Criterion C	12,071	42,042	28.71%	2.8	5	0.84
01 Criterion D	12,071	42,042	28.71%	3.2	5	0.79
02 Criterion A	29,848	42,042	71.00%	2.8	5	0.98
02 Criterion B	29,848	42,042	71.00%	2.7	5	0.93
02 Criterion C	29,848	42,042	71.00%	2.9	5	0.83
02 Criterion D	29,848	42,042	71.00%	3.2	5	0.79

Correlation between item and rest of component

Question	Correlation coefficient
01 Criterion A	0.8044
01 Criterion B	0.8018
01 Criterion C	0.7983
01 Criterion D	0.7253
02 Criterion A	0.7954
02 Criterion B	0.7944
02 Criterion C	0.7865
02 Criterion D	0.7320



Reviewing script evidence

Next, senior examiners compare the grade descriptors for each grade with scripts that have marks around the anticipated grade boundaries. During this phase of the grade award process, it is essential that examiners focus not on the marks awarded, but on the nature of the student responses and how well these match the grade descriptors.

Before starting the script review, it is helpful for examiners to look at examples of work from previous years to remind them of the expectations for different grades. Such scripts must be available during the grade award process.

Before the grade award meeting, members of the senior examining team—particularly the principal examiners—submit provisional grade boundaries for each component. They indicate at which marks they feel the boundaries should lie, based on their past experience of the expected standards. These provisional boundaries—together with the consensus on how each examination has functioned, and an awareness of the overall distribution of marks—provide the IB subject manager with the information and evidence they need in order to suggest a range of marks to be represented in a grade awarding sample.

Each senior examiner should review the selected student scripts and determine which grade descriptor best reflects the quality of work. This is a challenging task, as even students who have a relatively even level of response will often show qualities of a wide range of grades across their answers. It is acceptable, and often helpful, to indicate whether a particular script is just within a grade descriptor or almost reaching the next grade up (often indicated, for example, by a "7–" or a "6+").

Examiners should minimize any possible preconceptions or bias when undertaking this task. For this reason, it is important that they do not discuss their views with fellow examiners until everyone has recorded their independent result. Similarly, it is helpful for examiners (excluding the chief examiner) not to be informed of the findings of the statistical analysis until they have completed this stage, except for relevant information about individual item performance across the cohort.

Determining which grade is the best fit for a script is a very subjective exercise, and it is likely there will be variation between different examiners. Furthermore, it is not unreasonable for a student who has shown good understanding, but made a number of mistakes, to fit into a higher grade than a student who has excelled at the easiest task (and so gained many marks) but not shown the same depth of understanding, despite the latter student having slightly more marks. It is also important to remember that this script review only considers a relatively small number of examples of students' work.

Research has shown that examiners are most skilled at observing when scripts are of a different quality from others they are looking at (see the "Approaches to marking" section of this publication). The IB asks examiners to start at the highest mark in the sample set and to work down until they believe they have reached the point that they have stopped consistently seeing evidence of the higher grade. They should

then start at the lowest mark in the sample set and work up until they start consistently seeing evidence of the higher grade. If necessary, the range of marks in the sample set can be increased.

When all of the senior examiners' grading decisions are collected together, they should indicate the range of marks where a grade boundary should lie. This is called the "zone of uncertainty"; it represents the lowest to the highest marks where reasonable evidence exists to place the grade boundary. There is not a precise definition of how to set the zone of uncertainty—it should be agreed through a discussion among the senior examiners based on their individual decisions.

Figure 34

Example of script review outcomes

Script no	Mark	Examiner 1 (PE)	Examiner 2	Examiner 3	Examiner 4
1	51	4	4	4	4
2	51	4	4	4	4
3	50	3+	4	4	3+
4	50	4-	4	4	3+
5	49	4-	4-	4-	4-
6	49	3+	3	4-	3+
7	48	4-	3+	4-	3+
8	48	3+	4-	4-	3+
9	47	3	3	3+	3
10	47	3	3	3	4-

Zone of uncertainty

The only exception to this process is for multiple-choice question examinations. Experience shows that making judgements about grade boundaries based on the quality of work submitted for assessment is very difficult for examinations consisting only of multiple-choice questions. This may be because the responses contain very little evidence on which to base a judgement about what students have actually done. For such examinations, grade boundaries are calculated to fit, as far as possible, the same percentages of students within each grade as those established judgementally on the most closely associated examinations.

Reviewing statistics on outcomes

It could be argued that in a criterion-related system dependent on professional judgement, the senior examiners should be able to set grade boundaries purely by considering the questions on the examination and what each question requires from students in terms of a response. However, in reality it is very difficult to make these judgements without reference to how students have actually responded. Cresswell (2000) concludes that awarders are typically correct in identifying when examinations are easier or harder from one examination session to the next, but they are not typically correct in estimating how much easier or harder they are.

As the purpose of a grade award is to maintain a consistent meaning or standard for the grade, it is worth reflecting on one of Cresswell's definitions of comparability.

> Two examinations might be defined as having comparable standards if two groups of students with the same distribution of ability and prior achievement who attended similar schools with identical entry policies, are taught by equally competent teachers and are equally motivated receive grades which are identically distributed after studying the respective syllabuses and taking the examinations.
>
> (Cresswell, 1996, pp. 57–84)

To support the grade award, senior examiners are provided with the **statistically recommended boundaries**. These are defined as the grade boundaries that would give exactly the same cumulative percentage (the total proportion of students receiving the given grade or higher) as last year up to the given grade.

Examiners involved in the grade award process are also provided with information on the mean mark obtained by students, and with histograms of the actual distribution of marks. While this information is taken into account by the statistically recommended boundaries, it is often useful to be able to consider this greater level of detail.

It is important to remember that it will be a different group of students taking examinations each session. Therefore, it is likely that the assumptions of the definition of comparability given above will not be completely met. This is particularly true if the two cohorts are very dissimilar to each other. Equally, when dealing with large numbers of students from similar educational experiences (such as IB World Schools), it is more likely that significant differences in outcomes are due to those students answering a different set of questions, rather than variation in their performance.

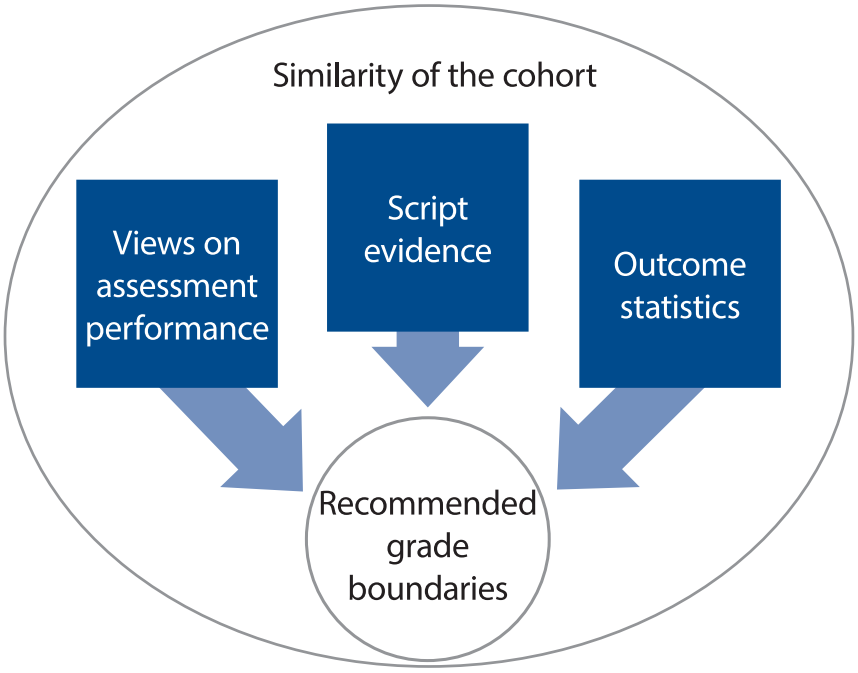
Cohort size must be taken into account when considering the significance of the statistically recommended boundaries, but the IB does not have formal rules around this as other factors are also important. In general terms, if there are only tens of students, considerable variation in the overall outcome might be expected; if there are thousands of students, this level of variation is relatively unlikely.

Balancing the evidence

No one type of evidence is more important than any other, and the task of the grade awarding team is to balance them all together in making their recommendation.

Figure 35

Evidence that supports the selection of grade boundaries



Often, the different evidence suggests the same outcomes—such as when the statistically recommended boundaries lie in the zone of uncertainty—but sometimes there is a contradiction between them. In such cases, it is essential to explain the discrepancy when justifying the final decision.

At this final stage of the grade award process, it is possible to model what effect different decisions would have on the overall outcomes for the cohort. Recalling that it is the overall result that is important, not the results for individual components, the awarding team might compare how the cohort's mean grade compares with teachers' predictions, referring to previous years to understand how reliable such comparisons have been. Another approach is the exclusion of new schools from the overall results to see if they are performing very differently to more established institutions.

Finally, the awarding team needs to submit their recommendations for approval, together with the justification for these recommendations. Aggregation then takes place. This is the process of combining

marks and boundaries from the different components to form a final mark or overall grade boundary. To achieve this, overall component marks or boundaries may need to be scaled. Please see the "Aggregation" section of this publication for more information.

Fixed grade boundaries

For many internally assessed tasks, and for some externally marked components, the task that the students are asked to complete is essentially the same for every session. Examples might include the preparation of an art portfolio or a personal project. In such cases, it is reasonable to assume that, as the IB maintains the same marking standard every year (through standardization with previous years' work), the grade boundaries will also remain consistent.

During the grade award process, the principal examiner and chief examiner will be asked to consider whether there is any evidence that the existing grade boundaries are not appropriate; in most cases, the IB would expect them to conclude that the existing boundaries should be carried forward.

Despite this, it is important to be clear that these boundaries are considered every year, and they can be adjusted if there is evidence that they are no longer effective in maintaining the comparability of grades between sessions. Examples of how this could occur include when there is a significant change in students' behaviour (such as new developments in technology for completing a task), changes in the work submitted that increase alignment with the markscheme without increasing the overall quality of the work, or a revision to the approach the IB takes in marking or moderating work that results in a change of marking standard.

For more information on why grade boundaries change, and what it means if they do, see "Setting grade boundaries: Why grade boundaries can change, and what it means if they do" (video).

Review of grade boundary recommendation and approval

Once the grade awarding team has reached a recommendation for where the grade boundaries should be placed, this recommendation and its justification are presented in the **24-hour distribution subject report**. This report also contains the underlying data, such as changes in the cohort taking the course and grade outcomes from the proposed boundaries compared with previous years. The report is then reviewed by senior members of the IB Assessment division to determine whether the arguments are sufficiently robust.

Where the IB Assessment leadership team is not satisfied with the recommendations made, they will discuss their concerns with the subject manager and ask them to place greater emphasis on one of the aspects of evidence or to provide more analysis to support their recommendations.

The IB's chief assessment officer has the final authority on where grade boundaries are placed, based on the recommendations from the grade awarding team.

Awarding a programme certificate

The IB programme certificates (MYP Certificate, IB Diploma, IB CP Certificate) are not awarded through a grade award process. They are determined during preparation for publication of results, and the process is described in the section "Preparation for release of results".

Teacher observers

The IB is committed to increasing the transparency of its assessment processes and increasing general understanding of how grades are awarded. As part of this commitment, teacher observers are invited to attend grade award meetings (precise details will depend on whether the meeting is face to face or being held virtually). There is an expectation that teacher observers will report back to colleagues on their

experience and provide a report for the wider IB educator network (IBEN) community. For more details, please contact IB Answers.

Principles of the grade award process

The underlying principles of the IB grade award process are as follows.

- The 3/4, 6/7 and 2/3 grade boundaries are determined (in that order) using all the available evidence (judgemental and statistical). Where there is no work submitted for assessment to establish these grade boundaries, the evidence that is available will be used to establish whichever boundary is most appropriate.
- The other grade boundaries are then determined arithmetically according to the appropriate procedure.
- Grade boundary decisions are made based on a triangulation of evidence from examiner judgement, statistical evidence and cohort information. All of these must be balanced equally and a compromise established.
- If the cohort for an assessment is broadly similar to previous years, then the IB would expect the outcomes to be broadly similar to previous years. However:
    - cohorts often do vary between years, particularly in small-entry subjects
    - where the outcomes do vary, the IB would expect strong evidence to understand why
    - if the tasks of an assessment are broadly similar to that of previous years, the IB would expect the grade boundaries to be similar to previous years
    - all grade boundaries can change between years, even for IA tasks
    - while the IB makes every effort to ensure consistency in difficulty of assessments between years, it recognizes that the demand of particular examinations will vary
    - the overall course grade boundaries are the priority; it is this outcome that is significant to the student and is used by stakeholders to make decisions
    - component grade boundaries are a key step to arriving at a robust overall course result, but small effects at component level can combine to have a large effect on the whole course outcomes.

Quality checks

- The most important outcome of assessment is the grade that students receive, so the final checks focus on this.
- The purpose of these final checks is to look for anything unusual.

After grade boundaries have been set, some additional quality checks are carried out to ensure the grades issued to students are fair and correct.

One of these checks is "at risk" re-marking. This identifies students with a final grade that is unexpectedly low. Some of the "at risk" re-marking concentrates on the work of examiners whose marking has been identified as needing further investigation. The re-marking is carried out by examiners who have proved they can consistently apply the principal examiner's marking standard. "At risk" re-marking is carried out not to identify whether extra marks can be found, but as a check that the current mark is appropriate, to ensure that students receive the correct grades.

IB Assessment staff also carry out a check to verify that subject grades for schools look appropriate.

IB Final Award Committee

- The IB Final Award Committee provides a formal setting where the IB outlines the assessment session for scrutiny by an invited range of stakeholders in relation to the MYP, DP and CP. The PYP does not have an IB Final Award Committee as there are no final results to verify or certify.
- The IB Final Award Committee consists of representation from across the organization, including senior IB staff and chief examiners. As well as giving final approval for the issue of results, it also considers cases against the IB's academic integrity policy, and any other pertinent issue.

The last stage in the approval of students' results for IB programmes with IB-submitted assessments is the IB Final Award Committee. It is this body that reviews the recommendation of the chief education officer and director of assessment that the session has met IB standards, and that results should be awarded on behalf of the IB Board of Governors. It is also a forum for setting policy and precedents relating to the award of IB qualifications.

Its precise remit and composition varies slightly between the different programmes, but in general it consists of equal numbers of voting members drawn from:

- senior IB staff covering the Education Office and Schools Office
- chief examiners from a range of different subject areas.

While senior assessment colleagues are responsible for approving grade boundaries, the IB Final Award Committee acts as an oversight board and reviews the macro-level outcomes, such as overall programme completion rates and any issues brought to its attention by colleagues in the IB Assessment division.

The committee also reviews issues of academic misconduct and maladministration, which are discussed at length by the academic integrity subcommittee.

The final role of the IB Final Award Committee is to reflect on the performance of the examination session and make recommendations to the IB for subsequent sessions.

Conflict of interest

The IB Final Award Committee is an important decision-making body, and it is critical that it is seen to be both transparent and independent. The policy is that any member who could be perceived as having a conflict of interest on a particular agenda item will leave the room for that discussion—this is reiterated at the start of every meeting.

Observers

The IB encourages observers to attend the IB Final Award Committee meetings in order to:

- make the procedures more transparent
- provide an opportunity for suggestions for change and improvements to the process
- acknowledge the partnership between IB constituents.

In order to support the IB in meeting these objectives, observers are requested to submit a report on their attendance to the chief education officer within two weeks of the meeting. This report should include general observations on current procedures and process, and suggested changes and improvements, where appropriate. No comments should be made on individual cases that were considered by the committee.

Owing to the sensitive nature of the issues discussed by the committee, observers are also bound by the appropriate restriction on confidentiality and personal issues of conflict of interest. Observers have no voting rights on the decisions of the committee.

For more details on being an observer at an IB Final Award Committee meeting, please contact support@ibo.org.

Preparation for release of results

- Once all assessments have been marked and grade boundaries determined, several processes need to be completed to ensure that students' results are ready to be published.
- In certain circumstances, where there is no evidence of work submitted for assessment, the IB may estimate a mark using the missing mark procedure.
- Subject results for students enrolled in a full programme are combined to determine whether an MYP Certificate, IB CP Certificate or IB Diploma has been achieved.
- The publication of results includes a robust change-control system. Any changes to grades are identified and the relevant stakeholders—primarily schools and universities—are informed.

Between setting grade boundaries and issuing results to students, a number of processes and procedures need to be completed. Some of these processes, such as carrying forward anticipated subjects, only apply to certain students, and others may affect all students.

Missing marks

- Where the IB is unable to access the student's work through no fault of the school or student, it will estimate a mark to minimize the potential disadvantage to the student.
- Such an estimation must be based on evidence. If there is very little work submitted for assessment on which to base an estimate, an alternative solution to the missing mark must be found.
- The missing mark procedure is based on average student performance; therefore, roughly equal numbers of students will be advantaged by its estimation as are disadvantaged. The fairest result is always to have actual work submitted for assessment to mark.

There are sometimes cases where a student's work is not available to the IB to mark for reasons beyond the control of the school and the student. Examples include when paper-based examinations are lost in the post or if students have a sudden illness on the day of the examination.

The missing mark procedure is a mechanism that the IB can use in such circumstances to estimate a mark for the student.

It is appropriate to invoke this procedure in circumstances where the student's work being missing is the result of the actions of the IB or third parties (not including the school), where it would not be reasonable for the student to be asked to complete the assessment on another occasion.

In all cases, the missing mark procedure must be based on evidence of the student's achievement. If the IB does not have much information on the student's performance, it will be very unlikely that a fair estimated mark can be given.

The missing mark procedure is based on the average attainment of all students in a given assessment component compared with their performance in other components. By taking this average, the IB would expect as many students to perform worse by this estimation as would perform better. It is, therefore, always a fairer result if the IB can mark the student's actual work.

If there are five or more students registered for the subject and level in a school, the IB uses the school's data to determine missing marks. If there are fewer than five students in a school, the IB uses global data to calculate missing marks.

Figure 36

Determining a student's missing mark

To determine the candidate's missing mark:

$$\text{Ratio} = \frac{\text{Candidate's total scaled mark on other components}}{\text{Global (school) total scaled mark on other components}}$$

$$\text{Candidate's missing mark} = \text{Global (school) scaled mark on missing components} \times \text{Ratio}$$

The ratio compares the student's marks with the global or school average for the components they have completed.

- If the student has done better than average, this number will be greater than one.
- If the student has done worse than average, this number will be less than one.

This calculation, like all missing mark procedures, is only the "best guess" based on the available evidence. It is always preferable to have the students' actual work to mark.

Missing grade procedure

For the MYP, if there is only one component in a particular subject's assessment model then the missing mark procedure cannot be used as there is no evidence from which to estimate a mark. In this situation, the IB uses a different approach: the MYP missing grade procedure.

One disadvantage of managing student burden by only having one assessment component is that there is only limited evidence available in the MYP. This means that the IB's confidence in the missing grade procedure is lower than in the missing mark procedure for other programmes; therefore, the procedure should only be used in exceptional cases.

Where a student has completed the assessment as required and submitted it in good faith, but it is not available for the IB to mark due to factors beyond the control of the student or school, the IB will apply the missing grade procedure.

In order to apply the missing grade procedure, the student must have been awarded a final grade in at least four other courses within the MYP. The IB is unable to estimate results for students with fewer results because it does not have enough evidence to make an informed estimation.

The missing grade calculation involves determining the mean grade from all other subjects with a grade determined by work submitted for assessment in the past 18 months.

If the mean grade is 0.5 or higher above a whole number—round up.

If the mean grade is less than 0.5 above a whole number—round down.

This calculation, like all missing mark procedures, is only the "best guess" based on the available evidence. It is always preferable to have the students' actual work to mark.

Programme outcomes

- The IB awards diplomas and certificates on the basis of meeting criteria, not on individual judgements.
- Each programme has its own criteria for the award of the diploma or certificate.

Once the IB has all subject results for a student, it can calculate whether they have qualified for the MYP Certificate, IB Diploma or IB CP Certificate.

Details about the how the MYP, DP and CP outcomes are calculated, and the passing criteria, can be found on the IB website and in the relevant programme's *Assessment procedures*.

Publication of results

From the perspective of the IB, the publication of results is primarily about implementing a very strict change-control protocol.

During the examination session, information is constantly updated so that the IB has a clear picture of what is going on. Once students and schools have been informed of their outcomes, the IB needs to be very clear that they do not make a change without everyone being properly informed. For the DP and CP, this particularly applies to universities, which receive transcripts of students' outcomes.

Authorized changes to results after they have been issued may occur due to enquiries upon results, following confirmation of pending results from schools or the resolution of academic integrity cases.

Enquiries upon results and assessment appeals

- The purpose of the enquiries upon results service is to allow schools to highlight to the IB where they believe a mistake has been made in the marking process.
- The same standards must be applied during enquiries upon results as during the main examination session, and the IB uses its seeding quality model to ensure this happens.
- External assessments are intended as summative rather than formative assessments, so the IB requires its examiners to only write comments when it supports their marking.
- The IB also has a formal appeals process if schools do not believe that the correct processes have been followed.

For the legal and procedural description of the enquiry upon results services offered, please refer to the relevant programme's *Assessment procedures*.

Categories of enquiries upon results

After the issue of results, the coordinator may request one of the following.

- **Category 1 enquiry upon results**—re-marking of all a student's externally assessed components for a subject
- **Category 1 report**—report on the marking of a category 1 enquiry upon results
- **Category 2 enquiry upon results**—copies of externally assessed component material
- **Category 3 enquiry upon results**—re-moderation of an internally assessed component

A fee is payable for each of the above categories (except when a grade is changed during a category 1 re-mark).

Why offer enquiries upon results

The enquiry upon results process is intended as a final safeguard against errors in the marking system. It is an opportunity for schools to highlight to the IB when they believe a mistake has been made, and for the IB to investigate and correct such errors, where necessary.

Maintaining standards from live marking to enquiry upon results marking

The purpose of enquiries upon results is to provide students with the marks that reflect the quality of their work based on IB standards. It is very important to ensure that examiners are not inappropriately influenced by the fact that the student is unhappy with their initial mark or is close to a grade boundary.

In the enquiry upon results process, the IB only uses its most consistent and senior examiners; it therefore regards the result of this re-mark as the correct mark.

To support examiners in maintaining the standard, the IB includes seed scripts within the enquiry upon results work that alert examiners electronically if they are moving away from the agreed standard.

IB assessment staff review every proposed enquiry upon results grade change to ensure that the new grade is more reliable than the original grade. When there is a large and unexplained mark difference between the two examiners, and a grade change, a third opinion on the script may be requested.

Category 1 enquiry upon results: Re-mark

As explained in the "Marking" section, two examiners could award different marks for the same question which are within an acceptable difference (tolerance). Therefore, if the enquiry upon results process results in only a very small change in the number of marks, then it is likely that there has not been a mistake in the marking. This is the case for the majority of the IB's enquiry upon results requests.

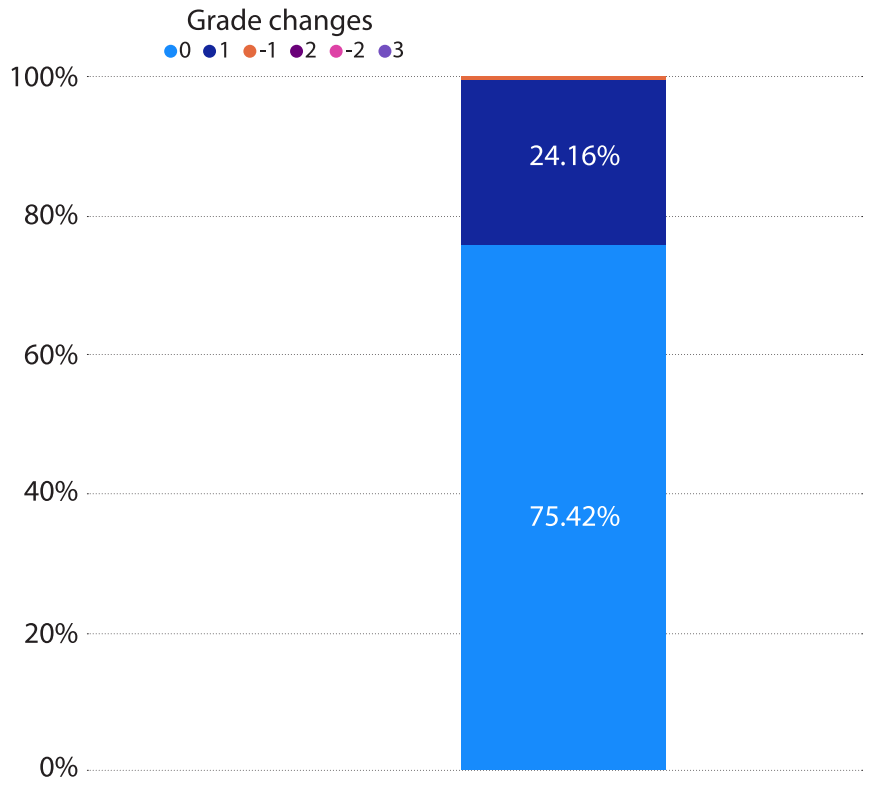
To minimize the impact of these small—and acceptable—differences in examiners' marks, the IB re-marks all the external assessments taken by a student during a category 1 enquiry upon results. This mitigates the impact of a small change as, assuming that any changes are not the result of systematic issues, the IB would expect them to cancel out over several examinations.

The IB believes that the enquiry upon results mark represents the most accurate representation of the student's work, and so this is the final mark awarded. This means that a student's result can go down as well as up as a result of a category 1 enquiry upon results. Schools must, therefore, obtain the student's permission before submitting a category 1 enquiry upon results.

Given the nature of setting grade boundaries, there will always be students who are 1 mark below the boundary, or just over it. In these cases, an enquiry upon results outcome could result in a change in grade despite there being no mistake in the marking. In cases such as these, the IB would emphasize that the student is on the boundary between two grades and either is a fair representation of their performance.

Figure 37

Average category 1 enquiry upon results grade changes over five DP examination sessions



Category 2 enquiry upon results: Return of student materials

The assessments that are considered for a category 2 enquiry upon results are summative assessments. 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 students' work.

The IB expects examiners to mark appropriately, to ensure that students get fair outcomes that reflect the work they have submitted. They are only asked to include comments that support their marking, not to provide feedback for students on how they could improve.

For more information on what to do if you think work has not been marked fairly, see "Challenging results: What you can do if you think your work hasn't been marked fairly" (video).

Category 3 enquiry upon results: Re-moderation

For a category 3 enquiry upon results (re-moderation), any change in marks will apply to the entire cohort. If the IB applied the principle of seeking students' permission, this would mean that a single student could stop their cohort's work from being re-moderated by not giving their approval or simply not responding to the request. This would clearly be unreasonable and so, to prevent students' grades being reduced without their formal acceptance of this risk, grades can **only increase**, not decrease, as a result of a category 3 enquiry upon results.

A re-moderation will only be undertaken in cases where the mean of the students' moderated IA marks differs from the mean of their raw marks (the marks awarded by the teacher) by **at least 15%** of the maximum mark for the component.

This threshold is in place because if the IB imposes no restriction, some schools will routinely ask for re-moderation in all their subjects in the hope that some grades will be increased. The IB's limited number of senior moderators would not be able to manage this level of demand. Setting this parameter for accepting re-moderation ensures that only those schools that have a clear concern can access the service. This allows the IB to manage the workload for its small number of senior moderators and, at the same time, offer re-moderation to schools where the effect of moderation has been severe on students' marks.

Figure 38

A cohort that qualifies for a category 3 enquiry upon results

Student number	Sampled	Raw marks	Moderated marks	Moderation adjustment
2	Y	19	14	-5
10		15	12	-3
11	Y	16	13	-3
12		20	15	-5
13		18	14	-4
16	Y	12	11	-1
36	Y	17	13	-4
45	Y	10	10	0
	Average change as a mark			-3.125
	Average change as a percentage of the max mark			-15.63%

In the example provided in figure 38, the maximum mark available for the component is 20.

The mean moderation adjustment is –3.125 marks, which is equivalent to 15.63% of the maximum mark.

This means that the sample qualifies for a category 3 re-moderation.

Assessment appeals

An assessment appeal is an opportunity to ask the IB to review how it made a decision as part of the assessment process, that is, the processing of an IB grade for a student or the handling of a case of student academic misconduct.

An appeal cannot ask for an exception to an IB process, but it can challenge the reasonableness of the interpretation of IB rules. An appeal cannot change the IB's published policies, and where an appeal results in a clarification of IB rules, it must be applied to all students in that session. It represents the final step in concerns about the IB assessment process.

For further information, please refer to the *Assessment procedures* publication for the MYP, DP or CP, available on the PRC.

Retake students

If a student is not satisfied with the grade they have achieved, they may take the subject's examinations again. This can be in the examination session six months later or, in principle, after any later examination session. However, when a subject is taken again, if significant new curriculum or assessment requirements have been introduced, the student must comply with those new requirements. A student who wishes to take one or more subjects again does not have to register in the same school where the subject was originally taken. If a school accepts a student who wishes to retake one or more subjects, the school must assume all academic and administrative responsibilities for that student.

A student taking a subject again can carry forward their result for a non-examination component, assuming there have been no significant changes to the curriculum and/or assessment requirements. A mark cannot be carried over from a written examination so a student would need to retake all examinations for the subject they are retaking.

If a student wishes to resubmit work for any non-examination component, the student must attend classes at the school where they are registered for the retake session. This is because the subject teacher must provide academic guidance, mark work for IA and confirm that all work is authentic. It is not sufficient for the student to make minor changes to non-examination work previously submitted. In principle, entirely different work should be submitted for assessment. However, the IB recognizes that for very extensive coursework tasks this may not be feasible given the time constraints. In such cases, substantial modification of the original work is acceptable. If modified work is submitted, it will be marked as new; the student must accept that the work may attract a lower mark. In the case of modified work, the student must have spent additional time equal to at least 20% of the total hours spent on the work. In addition, the work must meet at least one of the following criteria. It must have:

- a changed research question, focus or thesis
- a new theme, evidence strand or point of view that impacts the conclusion
- had substantive additional analysis/calculation
- explored a new (additional) aspect of an author or artist's work.

Where the student has work that has already been completed, but that has not been submitted as part of their assessment (for example, DP arts portfolios), it is acceptable to count time already taken to create this previously unseen work as contributing towards the 20% target.

This approach is selected to balance the expectation that retake students should not benefit from another round of feedback on work (to the disadvantage of students who present new work) while recognizing that it would not be fair to place an undue burden on students or the school to obtain, for example, entirely new data or research that would effectively prevent a resit.

The IB's approach to comparability can be found in the "Comparability" section of this publication.

The IB recognizes that over time the content of a course will need to be changed. Prior to introducing a new course, the IB communicates any changes to schools in various ways, including through the relevant *Programme updates* and the production of a new subject guide. After the last examination session of the previous course, the IB is unable to offer retake students an opportunity to sit examinations relating to the

previous course. This is partly because it would be unmanageable for both schools and the IB to have students being assessed on two different courses within the same subject in the same examination session. However, it may be possible for retake students to carry forward their IA mark or externally assessed coursework mark. If the new course is substantially different from the previous course, this option is unlikely to be available.

For more information regarding retake students, refer to the relevant programme's *Assessment procedures* publication.

Feedback to schools

- The purpose of IB summative assessment is to measure a student's performance, and all relevant processes are designed to maximize the validity of this outcome.
- The IB prioritizes transparency in communicating to schools and students about assessment results, so that students and teachers can see that marks and grades have been awarded consistently and without bias.
- Transparency is especially important with IA, in which the outcome is based on a moderation of the teacher marks.
- IB summative assessment, including the enquiry upon results process, is not intended to provide guidance to schools on how to improve student outcomes, except as a by-product of data required for a valid assessment.
- The IB does provide other services to schools to support their teaching pedagogy and professional development outside of the assessment process.

The purpose of IB summative assessment is to assess student performance following their course of study and provide a reliable measurement result in the form of a grade. This differs from formative assessment where the purpose is to provide feedback to assist a student with learning. While meaningful feedback is an essential part of formative assessment and effective teaching, it is not appropriate to use the outcomes of summative assessments for formative purposes. One reason for this is that an IB examiner will make their judgement on the one piece of work they have available—the examination script; in the classroom, IB teachers will draw upon a wide range of information and personal insight when deciding how to offer high-quality feedback to a student. Accordingly, examiners do not provide written feedback for students or teachers when marking.

Examiners are required to indicate clearly where marks have been awarded and to clarify with appropriate comments if there could be ambiguity. This supports the IB in checking standards and also provides transparency for schools on where marks have been awarded.

Subject reports

Following each examination session, the IB issues a subject report, usually produced by chief examiners. The subject report provides information about each assessment component and how students performed in the examinations, including questions or topics that were addressed particularly strongly or poorly. The report contains general recommendations from examining teams on how teachers may prepare students for similar questions or topics in future sessions. Subject reports also provide the grade boundaries for the subject and the individual component boundaries.

Individual teachers will naturally need to put information from the subject report into the context of their own classes. For example, while a particular question may have been poorly answered in general, it is quite possible that all their students obtained high marks.

Internal assessment feedback

Marks awarded by teachers for IA are subject to a moderation process. Teachers may interpret the IA assessment criteria differently from an IB examiner. For example, they may mark work more strictly and award fewer marks. The purpose of IA moderation is to adjust the teachers' marking standard to bring them in line with that of the principal examiner or the global standard. To help teachers understand how they are

varying from the global standard, the IB provides feedback on IA marking—when teachers' marks are found to be outside of tolerance—so that teachers can understand why a moderation factor has been applied.

The feedback provided to schools is not intended to explain how the students in the sample could have achieved a better result. Rather, it offers information on how well teachers marked each criterion, as well as their overall application of the assessment criteria and the suitability of the work submitted.

Note on textbooks, workshops and examinations

The curriculum of each IB course is set out in the subject guides, and this is the basis on which assessments are designed. Any textbooks, including those endorsed by the IB, are intended as aids to support students and teachers in completing the course as set out in the subject guides; they are not written to define the scope of the curriculum.

Therefore, if a topic, or part of a topic, is not included in a particular textbook but is in the guide, then questions may still be asked on it within the examinations. All teachers are, therefore, encouraged to refer to the subject guides exclusively when considering the scope of the curriculum. As part of the examination writing process, the IB does check commonly used textbooks to ensure that the examination questions do not duplicate those posed in any of these resources.

Similarly, comments and handouts from IB-run workshops do not replace statements set out in the subject guides, although they may support teachers in understanding how to interpret the wording in the guide. Any amendments to the guides will be formally published by the IB and clearly described as such.

Defining programme-specific processes

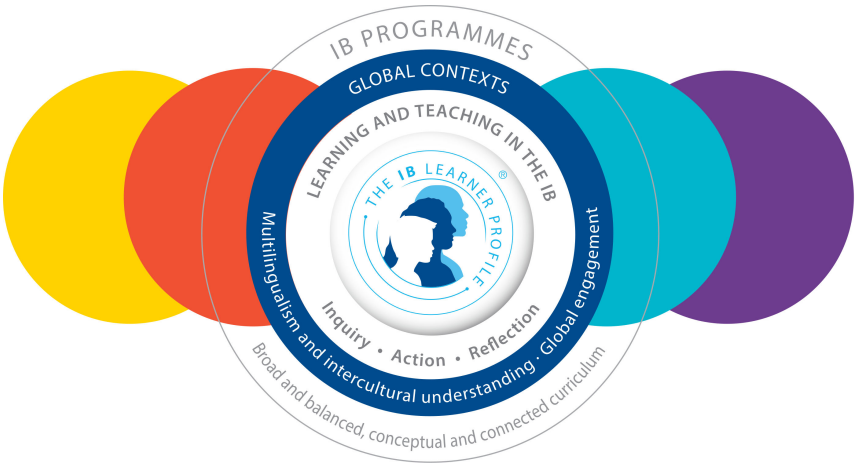
- As the same educational philosophy underpins all International Baccalaureate (IB) programmes, all IB assessments should follow the same principles and broad practices to meet this philosophy. The purposes of the assessment will naturally change as students progress through the different programmes at different stages of educational maturity.
- The broad purpose of external summative assessments in IB programmes is to provide students with a valid summary of their academic achievements to support them in progressing in education or work.
- While assessment is an important element of the Primary Years Programme (PYP), it is not appropriate for the programme to have any external summative assessment.
- Each programme has some distinctive features that require specific processes within the framework of the assessment practices.

Elements common to all programmes

- IB programmes offer curriculums or curriculum frameworks that are broad, balanced, conceptual and connected.
- All IB assessments need to consider these underlying aspects of an IB education in their design, even when they are not explicitly assessed, so that there is a positive backwash effect on learning and teaching.

Figure 39

Continuum between IB programmes



The publication *What is an IB education?* (2019) emphasizes that IB programmes promote conceptual learning, focusing on powerful organizing ideas that are relevant across subject areas and that help to integrate learning and add coherence to the curriculum. The programmes emphasize the importance of making connections, exploring the relationships between academic disciplines, and learning about the world in ways that reach beyond the scope of individual subjects. They offer students access to a range of academic studies and learning experiences that are broad, balanced, conceptual and connected.

- **Broad and balanced**—An IB education represents a balanced approach, offering students access to a broad range of content that spans academic subjects.
- **Conceptual**—Conceptual learning focuses on broad and powerful organizing ideas that have relevance within and across subject areas. They reach beyond national and cultural boundaries. Concepts help to integrate learning, add coherence to the curriculum, deepen disciplinary understanding, build the capacity to engage with complex ideas and allow the transfer of learning to new contexts.
- **Connected**—IB curriculum frameworks value concurrency of learning. Students encounter many subjects simultaneously throughout their programmes of study. They learn to draw connections and pursue rich understandings about the interrelationship of knowledge and experience across many fields. Course aims and programme requirements offer authentic opportunities to learn about the world in ways that reach beyond the scope of individual subjects.

When designing individual assessments, teachers, schools and IB authors need to reflect upon these underlying goals in order to avoid creating tasks that undermine good learning and teaching. The aim must always be to generate a positive backwash effect.

The international dimension

The mission of the IB is to help students develop as "caring young people who help to create a better and more peaceful world through intercultural understanding and respect" and "who understand that other people, with their differences, can also be right" (IB mission statement, 2024). IB programmes are also studied by students in many countries and regions (territories) and of many nationalities. There is, therefore, both an international context and a focus on developing international-mindedness in IB learning and teaching, both of which must be reflected in assessment. For further details on the IB's approach to international-mindedness, please refer to the section "International-mindedness and intercultural understanding" within this publication.

The IB learner profile

> Education today is much more about ways of thinking which involve creative and critical approaches to problem-solving and decision-making. It is also about ways of working, including communication and collaboration, as well as the tools they require, such as the capacity to recognize and exploit the potential of new technologies, or indeed, to avert their risks. And last but not least, education is about the capacity to live in a multi-faceted world as an active and engaged citizen. These citizens influence what they want to learn and how they want to learn it, and it is this that shapes the role of educators.
>
> (Schleicher, 2016)

There is a wide range of ways of categorizing these kinds of skills and competencies, including the Organisation for Economic Co-operation and Development's (OECD's) 21st-century competencies, RAND Education, the National Research Council (NRC) framework and others. Within the IB, these competencies are described through the learner profile.

Figure 40

The IB learner profile

Not all aspects of the learner profile are appropriate to measure through summative assessment. Good learning, teaching and assessment recognize the importance of these characteristics and—even when not designed to measure them—can offer students a chance to develop these competencies. Examples of this could be through encouraging ethical (principled) approaches to surveys and experimentation, supporting appropriate peer review and introducing unexpected contexts to students.

Further reading

For more information about the values that underpin the wider IB educational programme, please refer to the following resources.

- *What is an IB education?*
- The *IB learner profile*
- *From principles into practice* for the Middle Years Programme (MYP), Diploma Programme (DP) and Career-related Programme (CP)

Programme-specific needs and solutions

Earlier sections of this publication explain the processes that form the assessment cycle. These generalized practices hold true across all IB programmes. Each programme serves a different purpose, however, and so there will also be differences in some aspects of each programme's assessment, and in the rules to determine whether a programme certificate is given.

These approaches to assessment, which are specific to each programme, are important. This section explains what they are and sets out how they are managed.

Diploma Programme

The distinctive features of Diploma Programme (DP) external summative assessment are as follows.

- Students must take a prescribed set of subjects to achieve the diploma.
- Overall achievement in the diploma is described by a points score, with a maximum score of 45.
- Core components contribute up to 3 points to the overall diploma outcome via a points matrix.
- Nearly all subjects have multiple assessment components, including both external and internal assessment.
- Nearly all subjects are available at standard level (SL) or higher level (HL) and contribute equally to the overall diploma outcome.

Aims of the Diploma Programme

Assessment outcomes are only valid if they are clearly connected with the aims of the course and the programme.

The principal aim of the DP is to provide a challenging, internationally focused, broad and balanced educational experience for students aged 16 to 19. To support this aim, students are required to study six subjects and a curriculum core concurrently over two years. The DP also aims to equip students with the basic academic skills needed for university study, further education and their chosen profession, while supporting the development of the values and life skills needed to live a fulfilled and purposeful life. In other words, the validity of assessment outcomes in the DP, its courses and core elements is underpinned by concern for the whole educational experience of each student.

Valid uses for outcomes of DP assessments

When developing assessment models and curriculum, the IB intends that grades from DP courses and the overall diploma points score should be used to determine:

- selection for university admission or eligibility for employment
- whether a student has already met some of the requirements of a university programme, with potential acknowledgement of this by granting credit towards completion of a qualification, or being excused from taking particular courses
- evidence that a student is eligible to undertake further study in a particular language.

See the "Validity" section of this publication for why these valid uses are important considerations.

Structure of the DP and its courses

In order to achieve the IB Diploma, a student must take six subjects together with the core elements—theory of knowledge (TOK), the extended essay (EE), and creativity, activity, service (CAS).

Students will take some subjects at standard level (SL) and some at higher level (HL). Each student takes at least three (but not more than four) subjects at HL, and the remaining subjects at SL. SL and HL courses differ in scope, but performance on these courses is aligned through common subject-group grade descriptors and assessment objectives. Students are expected to demonstrate a greater degree of knowledge, understanding and skills at HL.

A limited number of interdisciplinary courses count across subject groups. For example, environmental systems and societies simultaneously satisfies the requirement to take a course from individuals and societies and sciences groups.

Figure 41

Diploma Programme model

How the DP outcome is calculated

The validity of the DP is reflected in the way a student's performance is measured. The overall diploma points score is calculated by adding together the grades (1 to 7) achieved for each of the six subjects, and then combining this with between 0 and 3 points for performance in the core elements (TOK, EE and CAS). SL and HL subjects are valued equally in determining a student's final point score. The highest score that a student can achieve is 45 points.

Core points matrix

Unlike the subjects, TOK and the EE are graded from A to E. The nature of the third element of the core, CAS, means that it is not assigned a grade.

Performance in the core contributes between 0 and 3 points towards the overall diploma points score. Obtaining a grade E in either TOK or EE, or not completing the requirements of CAS, will mean that a student is ineligible to receive the IB Diploma. The number of points awarded for the core is calculated as shown in figure 42.

Figure 42

Award of points for TOK and the EE

		Theory of knowledge (TOK)				
Extended essay	Grade awarded	A	B	C	D	E or N
	A	3	3	2	2	Ineligible for diploma
	B	3	2	2	1	Ineligible for diploma
	C	2	2	1	0	
	D	2	1	0	0	
	E or N	Ineligible for diploma				

Requirements for award of the IB Diploma

A student can receive the IB Diploma if the following conditions have been met.

- CAS requirements have been met.
- The student's total points are at least 24.
- An N (no grade awarded) has not been given for TOK, the EE or a contributing subject.
- At least a grade D has been awarded for both TOK and the EE.
- Grade 1 has not been awarded in any subject/level.
- Grade 2 has not been awarded three or more times (HL or SL).
- Grade 3 or below has not been awarded four or more times (HL or SL).
- The student has gained at least 12 points on HL subjects (for students who register for four HL subjects, the three highest grades count).
- The student has gained at least 9 points on SL subjects (students who register for two SL subjects must gain at least 5 points at SL).

Bilingual diplomas

As an alternative to the standard IB Diploma, a bilingual diploma certificate can be awarded to a student who fulfils **one or more** of the following criteria.

- Completes two languages selected from studies in language and literature with the award of a grade 3 or higher in both
- Completes one of the subjects from individuals and societies or sciences in a language that is not the same as the student's nominated studies in language and literature language. The student must attain a grade 3 or higher in both the studies in language and literature language and the subject from individuals and societies or sciences.

Pilot subjects and interdisciplinary subjects can contribute to the award of a bilingual diploma certificate, provided the above conditions are met.

The following cannot contribute to the award of a bilingual diploma certificate.

- An EE
- A school-based syllabus subject
- A subject taken by a student in addition to the six subjects for the IB Diploma (known as "additional subjects")

Managing standards across cohorts

Cohorts for subjects in the DP can range in size from a single student to over 40,000. While every subject is equally important, for practical reasons there are slightly different processes for subjects with large and small student numbers.

Maintaining examiner quality

It may not be appropriate to use the quality model of practice, qualification and seed scripts, described in the "Standardization" section of this publication, for subjects where there are very few examiners.

If all examiners are part of the discussion at standardization, then it is not necessary to provide practice scripts or qualification scripts, as every examiner has already been part of setting the standard. Seed scripts are still used to ensure that examiners continue to mark to the expected standard. However, a complete quality model is generally produced, even if it may not be needed.

The important consideration is that marking quality is maintained, even when there is a small number of students in a subject.

Grades: Full grade award (virtual or face to face)

For subjects where there are several examiners under the chief examiner or principal examiner, a formal meeting is held to follow the practices set out in the "Grade awarding and aggregation" section of this publication. An IB subject manager will be involved in the meeting to support the examiners and to provide a quality-control check on the process. These meetings may be face to face or virtual.

Grades: Guided small-entry subject grade award (virtual)

Where there is a very limited number of examiners in each component, the senior examiners will undertake the discussion virtually. The IB subject manager will provide them with the full range of evidence to support the award and monitor the process to ensure quality.

Grades: Standard small-entry grade awards

When the number of students is small enough that the statistical evidence between years is much less meaningful, examiners are asked to review all work submitted for assessment rather than just a sample at the judgemental grade boundaries.

IB subject managers support these meetings, as requested, and carry out random quality checks rather than routinely observing the meeting.

To support examiners in all small-entry subjects (guided and standard grade award), discussion between examiners in different subjects but within the same group is encouraged. This helps to ensure there is a common understanding of what grades mean across the group.

Final sign-off

Irrespective of the number of students who have taken the examinations, the final stage in the grade award process is for the chief examiner or principal examiner to make recommendations to IB senior assessment staff, who then confirm that they are convinced by the evidence supporting these recommendations.

Regardless of the number of students in a cohort, every subject is scrutinized in the same way.

Further reading

For more information about the DP, please refer to the following resources.

- *Diploma Programme: From principles into practice*
- *Diploma Programme Assessment procedures*
- *Rules for IB World Schools*
- Subject guides (see the subject pages on the Programme Resource Centre [PRC])
- Teacher support materials (see subject pages on the PRC)
- *Teaching and learning informed by assessment in the Diploma Programme*

Career-related Programme

The distinctive features of Career-related Programme (CP) assessment are as follows.

- Students must meet a set of requirements to achieve the IB CP Certificate.
- There is no overall points score associated with the IB CP Certificate.
- CP students taking courses shared with the Diploma Programme (DP) are assessed jointly with DP students.
- The CP framework requires a career-related study that is not offered or awarded by the IB.

Aims of the Career-related Programme

Assessment outcomes are only valid if they are clearly connected with the aims of the course and the programme.

The unique feature of the CP is that it supports students to become career-ready learners. The programme ensures that students develop the transferable and lifelong skills to support them throughout their further studies or employment, however they choose to progress.

The programme helps students to:

- develop a broad range of competencies and deepen their understanding in specific areas of knowledge through their DP courses and career-related studies (CRS)
- develop flexible strategies for knowledge and skill acquisition, or enhancement, in varied contexts
- foster attitudes and habits that empower them to become lifelong learners able to engage with diverse perspectives, opportunities and challenges
- prepare for effective participation in the changing world
- become involved in learning that develops their capacity and will to contribute in constructive ways.

Valid uses for outcomes of CP assessments

When developing assessment models and curriculums, it is intended that CP course grades and the IB CP Certificate can be used to determine:

- selection for employment and employment programmes, such as apprenticeships
- selection for further education in the appropriate vocational field of study
- selection for university
- whether a student has already met the requirements of a university programme (either through additional credit or being excused from taking particular studies/courses).

Where a student has taken assessments in a particular (response) language, that also provides evidence that they can undertake further study in that subject or vocation in that language.

See the "Validity" section of this publication for more information on why these valid uses are important considerations.

Structure of the CP

The CP is a three-part educational framework. It consists of:

- at least two courses from the DP at SL or HL
- the CP core components

- a career-related study.

Figure 43

Career-related Programme model

CP core components

The four components of the CP core are interrelated by their nature, learning outcomes and intentional connections. Viewed as a whole, the core aims to:

- anchor the programme to the IB mission through the development of the learner profile attributes and international-mindedness
- contextualize and enhance the DP courses and CRS, drawing together all the aspects of the framework
- promote the development of enduring personal, academic and professional knowledge, skills and attitudes
- reflect the understanding that learning is iterative, interconnected and a cornerstone of ongoing personal and community well-being.

Completing the four components of the CP core is mandatory. They are:

- community engagement
- language and cultural studies
- personal and professional skills
- reflective project.

The reflective project is assessed, meaning that the final project and reflection are marked and the student receives a final grade for the component from the IB. For the other core components, the school must confirm to the IB that they have been completed satisfactorily but the IB does not verify the assessment made by the school. Processes and practices employed by CP schools are reviewed during school evaluation visits.

DP courses as part of the CP

Each student takes at least two courses that are common to the DP. CP students are included in the same assessment process as DP students. There is no separate examination or grade award for CP students.

A student cannot be registered simultaneously for the DP and the CP, despite courses being common to both. The extent of the wider programme requirements of each preclude them being completed simultaneously.

Career-related study

The IB does not assess or apply any sort of quality control to the outcomes of the career-related study element of the CP. The only requirement is that the school confirms that the student has completed it. DP courses are not appropriate to form part of the career-related study portion of the CP.

How the CP outcome is calculated

There is no point score associated with the IB CP Certificate.

The IB CP Certificate will be awarded to a student provided all the following requirements have been met.

- The school has confirmed that the student has completed the specified career-related study.
- The student has been awarded a grade 3 or higher in at least two DP courses.
- The student has been awarded at least a grade D for the reflective project.
- The school has confirmed that all community engagement, language and cultural studies, and personal and professional skills requirements have been met.
- The student has not received a penalty for academic misconduct from the Final Award Committee.

The CP grades and outcomes are confirmed by the same Final Award Committee as the DP.

Bilingual CP certificates

As an alternative to the IB CP Certificate, a bilingual certificate for the CP can be awarded to a student who fulfils **one or more** of the following criteria.

- Completes two DP language courses selected from studies in language and literature with the award of a grade 3 or higher in both
- Completes a DP language course from studies in language and literature, and also completes a DP course from individuals and societies or sciences in a response language different from that taken in studies in language and literature. The student must attain a grade 3 or higher in both courses.

Managing standards and missing work

The reflective project and all courses that are common to the DP use the same approaches as the DP for the missing mark procedure and managing standards across the range of cohort sizes.

For more information on missing work and missing marks, see the "Missing marks" section of this publication.

Further reading

For more information about the CP, please refer to the following resources.

- *Career-related Programme: From principles into practice*
- *Career-related Programme Assessment procedures*
- *Overview of the Career-related Programme*
- *Reflective project guide*
- *Language and cultural studies guide*
- *Personal and professional skills guide*

- *Community engagement guide*

Middle Years Programme

The MYP Certificate recognizes that the student has achieved all aspects of the Middle Years Programme (MYP). To achieve the MYP Certificate, a student must:

- achieve a total of at least 28 points, with no grade 1 or grade 2 in any MYP eAssessment
- study the programme for preferably the recommended two years at least, with a minimum of one year, and complete requirements in year 5
- complete the internal assessments and examinations for a minimum of five disciplines from different subject groups and a maximum of eight disciplines from the eight subject groups, which must include the required subjects
- complete at least one disciplinary ePortfolio from the arts, physical and health education, or design groups
- complete the interdisciplinary on-screen examination
- complete and submit a personal project
- complete the IB's minimum requirements for community engagement.

Aims of the Middle Years Programme

Assessment outcomes are only valid if they are clearly connected with the aims of the course and the programme.

> The MYP has been designed as a coherent and comprehensive curriculum framework that provides academic challenge and develops the life skills of students from the ages of 11 to 16. These years are a critical period in the development of young people. Success in school is closely related to personal, social and emotional well-being. At a time when students are establishing their identity and building their self-esteem, the MYP can motivate students and help them to achieve success in school and in life beyond the classroom. The programme allows students to build on their personal strengths and to embrace challenges in subjects in which they might not excel. The MYP offers students opportunities to develop their potential, to explore their own learning preferences, to take appropriate risks, and to reflect on, and develop, a strong sense of personal identity.
>
> (*MYP: From principles into practice*, 2014, p. 3)

The MYP has an explicit alignment between the MYP subject group objectives and marking criteria. All MYP subject groups have four assessment criteria that match the four objectives. Each criterion contributes equally to the final outcome.

Inclusion of the global context in eAssessments

> In the MYP, learning contexts should be (or should model) authentic world settings, events and circumstances. Contexts for learning in the MYP are chosen from global contexts to encourage international-mindedness and global engagement within the programme … The MYP identifies six global contexts for teaching and learning that are developed from, and extend, the PYP's transdisciplinary themes.
>
> (*MYP: From principles into practice*, 2014, p. 18)

Each examination session will be shaped within, and informed by, a specific global context and exploration selected from the list published in *MYP: From principles into practice*.

Approximately one third of tasks within each disciplinary on-screen examination will be connected with, inspired by or derived from the selected global context. The whole of the interdisciplinary on-screen examination is inspired by the selected global context.

Partially completed unit planners for language acquisition, the arts, design, and physical and health education will be developed with reference to the selected global context.

Valid uses for outcomes of MYP assessments

When developing assessment models and curriculums, it is intended that grades from MYP courses and the certificate point score should be used for:

- selection for further educational opportunities or work
- positive feedback and an indication of personal strengths for students continuing their education.

Where a student has taken the assessment in a particular (response) language, that provides evidence that they can undertake further study in that subject in that language, for example, studying at a French-, Spanish- or English-speaking school.

See the "Validity" section of this publication for more information on why these are important considerations.

Structure of the MYP

Figure 44

Middle Years Programme model

How the MYP outcome is calculated

At the end of their year 5 MYP studies, students can be entered for the IB external assessment. The outcomes of these assessments will be recorded in an *MYP Course Results* publication. In addition, these students can choose to take assessments that can lead to the award of the MYP Certificate.

The school can also issue an MYP record of participation. This is for MYP students who study the programme for at least two years and complete the requirements in year 3 or year 4. These students are not

registered with the IB for any form of assessment. The MYP record of participation is a school-based document, not verified by the IB.

In order to achieve the MYP Certificate, a student must have participated in the final year of the programme, with a recommended participation period of two years, and:

- complete either an on-screen assessment or ePortfolio in six subjects, consisting of: language and literature; language acquisition (or a second language and literature); individuals and societies; mathematics; sciences; and one subject from the arts, physical and health education, or design
- achieve at least a grade 3 in each of the six subjects listed in the bullet above
- complete the on-screen interdisciplinary examination and achieve at least a grade 3
- complete the personal project with at least a grade 3
- obtain a total of 28 points overall
- meet the school's expectations for community engagement.

The MYP bilingual certificate additionally requires successful results from on-screen examinations for one of the following.

- A second language and literature course (instead of a course in language acquisition)
- One (or more) science, individuals and societies, or interdisciplinary examination(s) in a language other than the student's chosen language and literature course

Delivering external summative assessment (MYP eAssessment)

The optional eAssessment comprises two different ways to assess what students know and can do, including:

- ePortfolios of work submitted for assessment in language acquisition, the arts, design, and physical and health education, which are then moderated to ensure a consistent global standard
- on-screen examinations (two hours in duration) for courses in language and literature, language acquisition, individuals and societies, sciences, mathematics and interdisciplinary learning.

In addition, the personal project is submitted electronically to the IB and moderated. While other eAssessments are optional for schools, all MYP year 5 students must take part in the personal project eAssessment.

Examination blueprints

The IB publishes examination blueprints to provide clear guidance to schools on what the eAssessments will look like. These blueprints enable teachers and students to understand the nature and purpose of MYP eAssessment. They assist students in preparing for on-screen examinations and help students to focus on the subject-group criteria and assessment strategies in each subject group. There are always four criteria in the blueprint and each of these criteria is equally weighted.

The IB undertakes to ensure that, in any session, examinations will not deviate from the blueprint by more than 3 marks.

ePortfolios and partially completed unit planners

ePortfolios allow the assessment of an extended coursework task (product) or performance, which by their natures are difficult to test in an examination. The bases of the ePortfolios are the partially completed unit planners that guide the teacher in ensuring that appropriate student evidence is produced to allow fair and meaningful judgements to be made, as well as providing flexibility to meet local contexts. New partially completed unit planners are provided for each session.

The unit planners should ensure that the tasks set by teachers allow students to show evidence across the full range of MYP grades. There is clearly a risk with teacher-devised assessment that students are disadvantaged by an unreasonably easy or difficult set of tasks. In moderation, the IB can only award grades

based on the available work submitted for assessment. If the teacher-devised task only covers part of the range of grade descriptors, grades outside that range cannot be awarded.

Single assessment—managing student burden

Taking examinations and doing coursework is stressful and demanding for students. It can also take away from time spent teaching. For year 5 students studying the MYP, the IB believes that, on balance, it is more appropriate to minimize the amount of summative assessment. While this does create difficulties with students only having one opportunity to demonstrate what they can achieve, the IB accepts these problems to ensure the overall welfare of the students.

Approach to missing grades

One disadvantage of managing student burden by only having one assessment component is that there is only limited evidence available in the MYP. This means that the IB's confidence in the missing grade procedure is lower than in the missing mark procedure for other programmes and should, therefore, only be used in exceptional cases.

For detailed information on the missing grades process, please refer to the section "Missing grade procedure" of this publication.

Further reading

For more information about the MYP, please refer to the following resources.

- *MYP: From principles into practice*
- *Guide to the MYP exam session*
- MYP subject guides
- MYP projects guides
- *Middle Years Programme Assessment procedures*
- *MYP on-screen examinations: IT requirements and school responsibilities*
- *MYP on-screen familiarization*

Primary Years Programme

The distinctive features of Primary Years Programme (PYP) assessment are as follows.

- Assessment involves teachers and students collaborating to monitor, document, measure, report and adjust learning.
- Fostering an assessment culture involves the development of assessment capability among all members of the learning community.
- There is no requirement for IB external summative assessment.

Aims of the Primary Years Programme

Assessment outcomes are only valid if they are clearly connected with the aims of the course and the programme.

The PYP curriculum framework centres on transdisciplinary learning as the curriculum organizer for students to experience learning between, across and beyond traditional subject boundaries.

Informed by constructivist and social-constructivist learning theories, the PYP curriculum framework begins with the premise that PYP students are agents of their own learning and partners in the learning process. They have innate potential to inquire, question, wonder and theorize about themselves, others and the world around them.

This understanding of how students learn is foundational to the inquiry-based and concept-driven transdisciplinary model of learning and teaching. Through engaging with the programme of inquiry and reflecting on their learning, PYP learners develop knowledge, skills and attitudes to engage with the world —taking action for the well-being of people and the planet.

Structure of the PYP

Figure 45

Primary Years Programme model

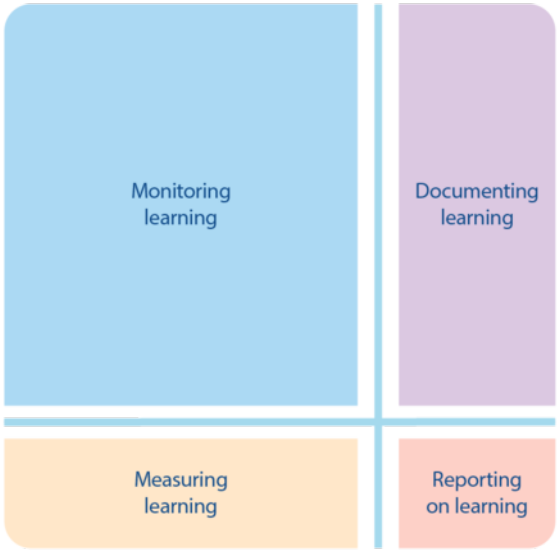
Assessment in the PYP

Assessment is central to the PYP goal of thoughtfully and effectively supporting students through the acquisition of subject-specific knowledge and skills, conceptual understanding, and the development of approaches to learning and the attributes of the IB learner profile.

PYP assessment has four dimensions: monitoring, documenting, measuring and reporting on learning. Each of these aspects has its own function, but all aim to provide evidence to inform learning and teaching. Although the four dimensions of assessment are not weighted the same, each dimension has its own importance and value. The PYP chooses to put emphasis on monitoring and documenting learning, as these dimensions are critical in providing actionable feedback for the learner.

Figure 46

Dimensions of assessment in the PYP



The following assumptions underpin assessment in the PYP.

- Assessment is an ongoing collaborative process between the teacher and student, and between peers, to gather, analyse, reflect and act on evidence of learning to inform teaching.
- A range of strategies should be used for assessing work, taking into account the diverse, complicated and sophisticated ways in which individual students understand their experiences.
- Students actively engage in assessing and reflecting on their learning, acting on feedback from peers and teachers to feed forward to the next steps in learning.
- Learning goals and success criteria are co-constructed and clearly communicated.
- Both learning outcomes and the learning process are assessed.
- Assessment design is both backward- and forward-looking.
- Fostering an assessment culture involves the development of assessment capability among all members of the learning community.

Further reading

For more information about the PYP, please refer to *PYP: From principles into practice* on the PRC.

Moderation of internal assessment: A closer look

Moderation is used with internally assessed work in the Middle Years Programme (MYP), Diploma Programme (DP) and Career-related Programme (CP) 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 their marking of the students' work.

Sampling

Once the school has submitted its internal assessment (IA) marks, the International Baccalaureate (IB) selects the students for the moderation sample. This sample is carefully selected to ensure that the mark range of the school is appropriately represented. Moderation sample sizes are ten, eight, five or fewer than five, according to the number of students in the subject cohort. Within the sample, three students will be allocated to the initial sample, to represent the sample mark range and determine whether a teacher is marking within tolerance of the global standard. If this is not the case, the rest of the sample (known as the "enhanced sample") will also be marked by the examiner. The IB tends not to select students for the moderation sample who have attained full marks, to allow students in the higher mark range the possibility of being moderated upwards.

When the school entry for a given course is large enough to split into different classes and more than one teacher is involved in carrying out the IA, the IB requires these teachers to share the IA and work together to ensure they have standardized between them the way in which they apply the criteria. A single moderation sample is requested from the school, which in all probability will contain work submitted for assessment marked by the different teachers involved. However, where there are different classes within one school using different response languages for the same subject, then a separate moderation sample is required for each language.

Determining moderation factors

All internally assessed components are marked by applying assessment criteria or markbands, and in most cases the teacher has access to considerably more information about the context and process underlying the work submitted for assessment than the examiner can have. Because of this, examiners moderating internally assessed components are asked to judge whether the teacher's marking seems appropriate, rather than simply to re-mark the work disregarding the marks awarded by the teacher. Teachers' marks should be altered only when the moderator is sure they are inappropriate and do not correspond to the global marking standard.

The teacher's marking sample is moderated by the examiner and then, based on a statistical comparison between the two sets of marks, where necessary an adjustment is made to the teacher's marks for all students 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.

In the following example (table 8), the teacher has under-marked their students by an average of 5 marks, so 5 marks will be added to all teacher marks in the cohort.

Table 8

Consistent under-marking

Teacher mark	Examiner mark	Difference between teacher and examiner marks	Final moderated mark
15	20	5	20
11	17	6	16
10	14	4	15
8	12	4	13
4	10	6	9

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 the next example (table 9), the teacher has marked inconsistently compared to the global standard. They have been too generous at the top of the mark range but too harsh at the bottom of the mark range, so the adjustment applied to the teacher marks will vary across the mark range.

Table 9

Inconsistent marking

Teacher mark	Examiner mark	Difference between teacher and examiner marks	Final moderated mark
37	35	–2	35
25	22	–3	24
17	18	1	17
12	14	2	13
6	8	2	8

If the teacher is marking to the correct standard, no adjustment will be made.

The teacher marks in table 10 are all within tolerance of the global standard, so all students will receive their teacher marks.

Table 10

Marking to the correct standard

Teacher mark	Examiner mark	Difference between teacher and examiner marks	Final moderated mark
32	31	–1	32
29	30	1	29
8	8	0	8

Linear regression

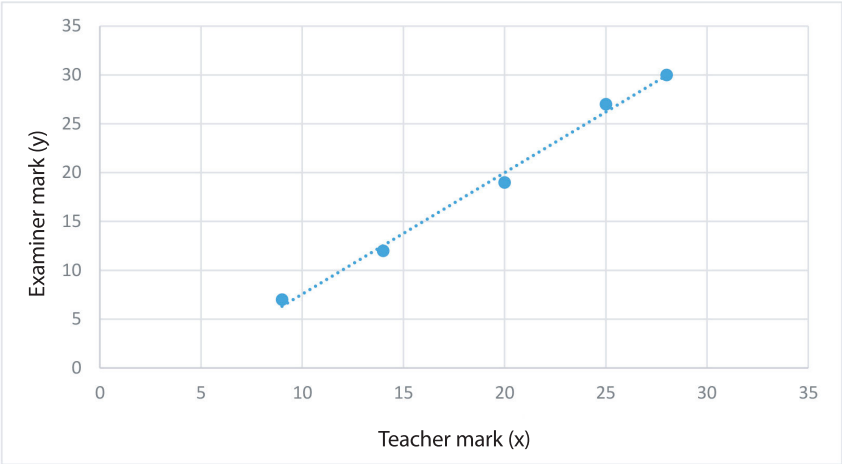
An analysis is carried out on the data for each moderation sample, which initially determines if a teacher's marking is within tolerance. If it is not, then an appropriate adjustment will be applied to all the teacher's marking based on the general trend shown in the sample. The technique used is called "linear regression",

which involves calculating the best-fitting straight line through the set of data points derived from the sample marks awarded by both the teacher and the examiner. An example of linear regression and the corresponding set of marks are shown in figure 47.

The moderation regression line in figure 47 shows a teacher who is slightly too harsh at the top end of the mark range and too generous at the bottom. Each individual point represents the pair of marks given to a piece of sample work by the teacher and the examiner. The continuous regression line is used to convert the teacher's marks into moderated marks.

Figure 47

Linear regression with the corresponding set of marks



Teacher mark	Examiner mark	Final moderated mark
28	30	30
25	27	26
20	19	20
14	12	12
9	7	6

The equation of the regression line calculated from the sample data can be used to convert each mark (x) awarded by the teacher into an equivalent mark (y) that the examiner would, on average, most probably have given to that same student. Such a moderation adjustment, based on extrapolating from a sample to a much larger collection of marks, can only reflect the general trends apparent in the marking. The purpose of moderation is to ensure that student marks, overall, are adjusted to more appropriate levels. It is a best-fit approach, and therefore not all students will receive the examiner's mark as their moderated mark.

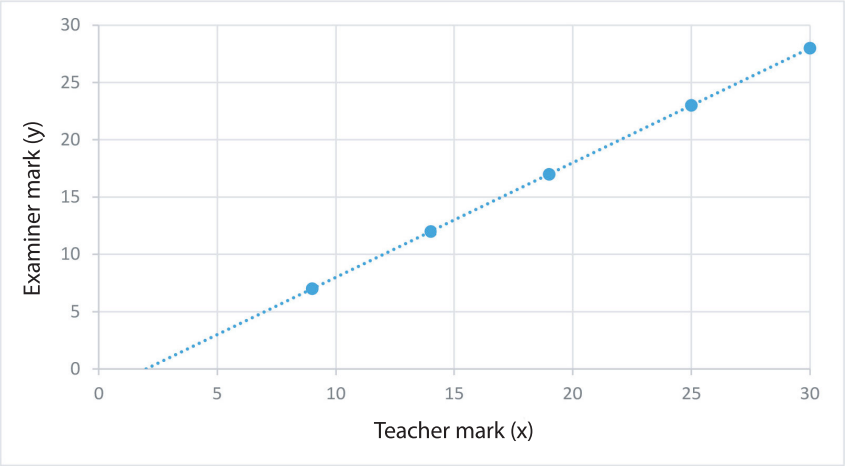
Moderation failure

A check is automatically carried out to make sure that the linear regression line of best fit (that is, the calculated moderation factor) meets certain conditions before it is applied to all a teacher's marks. In some cases, it may not be possible to calculate a moderation adjustment using the submitted sample work. One statistical measure is the correlation coefficient (the product moment correlation coefficient is used). This measures the consistency of the relationship between the teacher's and the examiner's marks.

A correlation coefficient of 1 indicates perfect consistency in the relationship between the marking and agreement in ranking students from best to worst (though not necessarily the same marks)—as shown in figure 48.

Figure 48

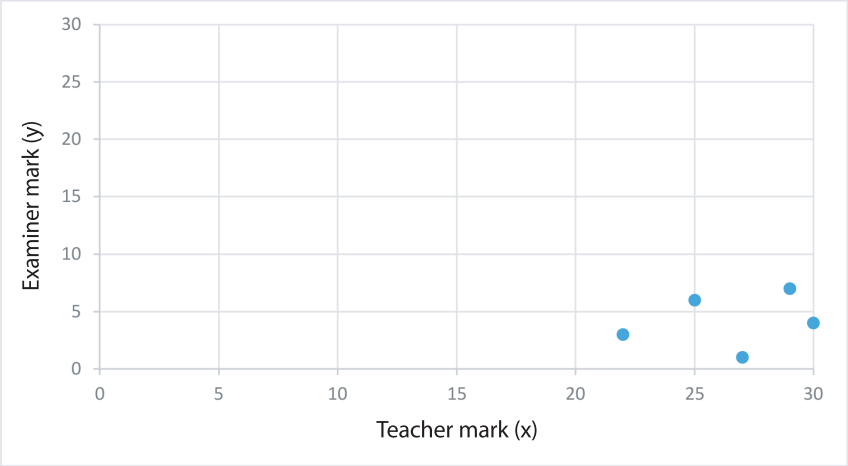
Correlation coefficient of 1



A coefficient of 0 indicates no relationship at all—as shown in figure 49.

Figure 49

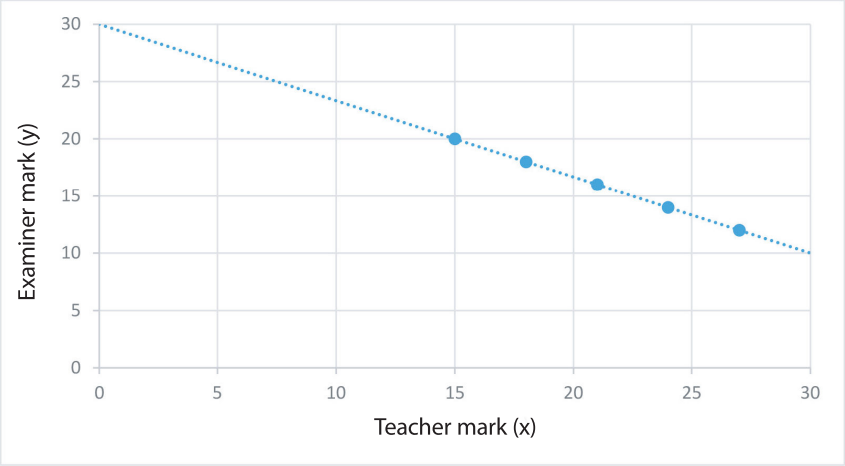
Correlation coefficient of 0



A coefficient of –1 indicates consistently opposing views regarding the relative merits of students' work, with the teacher and the examiner producing opposite rankings to each other—as shown in figure 50.

Figure 50

Correlation coefficient of –1



For the calculated moderation factor to be acceptable, the correlation coefficient must be at least 0.85, indicating a high level of agreement between the teacher and the examiner. However, a high correlation coefficient on its own is insufficient to ensure the moderation factor is appropriate. A further check is carried out that the gradient (slope) of the regression line is between 0.5 and 1.5. If the gradient of the line is too low (or too shallow), it means that the teacher has spread students' marks out too much, giving comparatively too few marks to weak work and too many marks to good work, even though this may be done on a consistent basis. The examiner has had to compress the teacher's mark range considerably. If the gradient is greater than 1.5, the line is too steep and the opposite applies (that is, the teacher has not differentiated sufficiently between poor and good work submitted for assessment, and the examiner has had to expand the mark range awarded).

A sample will fail the automatic moderation checks if the correlation coefficient is less than 0.85 or the gradient is outside of 0.5 to 1.5.

All cases of school samples that fail moderation are reviewed individually by IB assessment staff who consider the underlying data carefully and may decide:

- that the calculated regression line is appropriate for the teacher's particular mark range
- to apply some other moderation adjustment that is appropriate for the teacher's mark range
- to request further sample data in order to clarify the trend
- to request the rest of the students' work to carry out a complete re-mark of the teacher's marking.

All students' IA work must be available until the close of the examination session so that the IB can resolve any moderation failures.

Adaptation of linear model

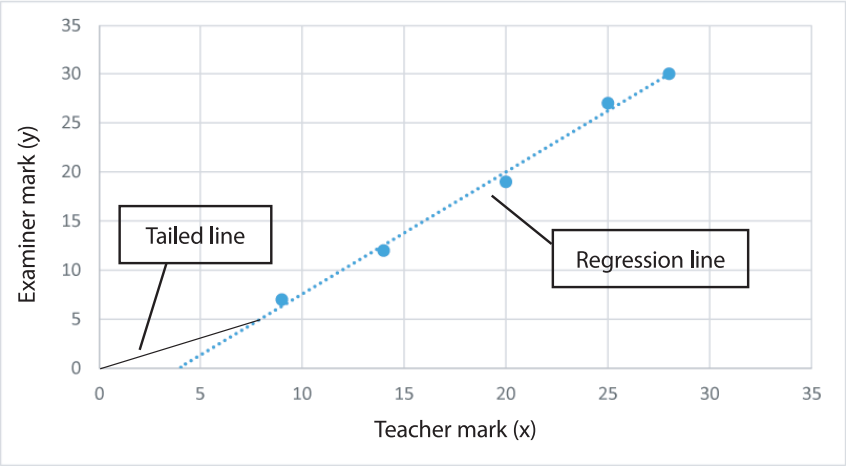
The straight-line model used for moderation is modified to some extent using "tailing". A straight-line moderation adjustment may have inappropriate effects at the extremes of the possible mark range, making it impossible for any student to be awarded zero. A student whose work is genuinely worth zero might be awarded a few marks when nothing of any worth had been written on the work, or a student whose work is poor but worth a few marks may be awarded zero through moderation.

To overcome this problem, "tailing" can be applied to marks in the bottom 20% of the available mark range. At this extreme, the calculated regression line is modified and substituted by a new "tailed" line that links from the regression line to the minimum coordinates. Tailing of the regression line is not applied at the upper end because it is common to see teachers awarding maximum marks to students who do not deserve them.

The graph in figure 51 shows the tailing of a regression line to prevent student marks from being adjusted towards minimum values.

Figure 51

Tailed regression line



Tailing ensures that a moderated 0 mark can only be derived from an original 0 mark. It prevents work that is worth a small number of marks from being awarded zero, and also prevents work that is worthless from being awarded a small number of marks. This assumes that the teacher's marking passes through the automatic moderation process. If the teacher's marking fails moderation and cannot be automatically moderated, tailing is not applied.

Bibliography

Ananiadou, K., & Claro, M. (2009). 21st century skills and competences for new millennium learners in OECD countries. *OECD Education Working Papers* (No. 41). OECD.

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Baird, J. A., Cresswell, M., & Newton, P. (2000). Would the real gold standard please step forward? *Research Papers in Education*, *15*(2), 213–229. https://doi.org/10.1080/026715200402506

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*(3), 347–364. https://www.jstor.org/stable/3448076

Black, P. (1999). Assessment, learning theories and testing systems. In P. Murphy (Ed.), *Learners, learning & assessment* (pp. 118–134). Paul Chapman Publishing in association with The Open University.

Bloom, B. S. (Ed.), Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: Handbook 1: Cognitive domain*. Longman.

Broadfoot, P. M. (1996). *Education, assessment and society: A sociological analysis*. Open University Press.

Brown, R. (2002). Cultural dimensions of national and international educational assessment. In M. Hayden, J. Thompson, & G. Walker (Eds.), *International education in practice: Dimensions for schools and international schools*. Routledge.

Chamberlain, S. (2010). *AQA: Public perceptions of reliability*. In Ofqual's *Reliability compendium* (Chapter 18). The Office of Qualifications and Examinations Regulation.

Coe, R., Searle, J., Barmby, P., Jones, K., & Higgins, S. (2008). *Relative difficulty of examinations in different subjects*. Report for SCORE (Science Community Supporting Education). CEM Centre, Durham University.

Cresswell, M. J. (1986). Examination grades: How many should there be? *British Educational Research Journal*, 12(1), 37–54.

Cresswell, M. J. (1996). Defining, setting and maintaining standards in curriculum-embedded examinations: Judgemental and statistical approaches. In H. Goldstein & T. Lewis (Eds.), *Assessment: Problems, developments and statistical issues* (pp. 57–84). Wiley.

Cresswell, M. J. (2000). The role of public examinations in defining and monitoring standards. *Proceedings of the British Academy*, *102*, 69–120.

Crooks, T. J., Kane, M. T., & Cohen, A. S. (1996). Threats to the valid use of assessments. *Assessment in Education: Principles, Policy & Practice*, *3*(3), 265–286. https://doi.org/10.1080/0969594960030302

Dolan, R. P., Burling, K., Harms, M., Strain-Seymour, E., Way, W., & Rose, D. (2013). *A universal design for learning-based framework for designing accessible technology-enhanced assessments* (Research Report). Pearson. https://eric.ed.gov/?id=ED576691

Frith, D. S., & Macintosh, H. G. (1984). *A teacher's guide to assessment*. Stanley Thornes.

Gibbs, G. (1992). *Teaching more students: Assessing more students*. The Oxford Centre for Staff Learning and Development, Oxford Brookes University.

Gipps, C., & Murphy, P. (1994). *A fair test? Assessment, achievement and equity*. Open University Press.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist, 18*(8), 519–521. https://doi.org/10.1037/h0049294

Goldstein, H. (1996). Group differences and bias in assessment. In H. Goldstein, & T. Lewis (Eds.), *Assessment: Problems, developments and statistical issues* (pp. 85–93). John Wiley & Sons Ltd.

Good, F. J., & Cresswell, M. J. (1988). *Grading the GCSE*. Secondary Examinations Council.

He, Q., Opposs, D., & Boyle, A. (2010). A quantitative investigation into public perceptions of reliability in examination results in England. In Ofqual's *Reliability compendium* (Chapter 19, p. 68). The Office of Qualifications and Examinations Regulation.

Hieronymus, A. N., & Hoover, H. D. (1986). *Iowa tests of basic skills: Manual for school administrators*. Riverside Publishing Company.

Hughes, D. C., Keeling, B., & Tuck, B. F. (1983). Effects of achievement expectations and handwriting quality on scoring essays. *Journal of Educational Measurement*, *20*(1), 65–70. https://www.jstor.org/stable/1434941

Humphreys, L. G. (1986). An analysis and evaluation of test and item bias in the prediction context. *Journal of Applied Psychology*, *71*(2), 327–333. https://doi.org/10.1037/0021-9010.71.2.327

International Baccalaureate. (2019). *What is an IB education?* International Baccalaureate Organization.

International Baccalaureate. (2024). IB mission statement. International Baccalaureate Organization.

Lambert, D., & Lines, D. (2000). *Understanding assessment: Purposes, perceptions, practice*. Routledge Falmer.

Linn, M. C. (1992). Gender differences in educational achievement. In J. Pfleiderer (Ed.), *Sex equity in educational opportunity, achievement, and testing*. Educational Testing Service.

Llewellyn, D. (2014). *Inquire within: Implementing inquiry- and argument-based science standards in grades 3–8* (3rd ed.). Corwin.

Meyer, A., Rose, D. H., & Gordon, D. (2014). *Universal design for learning: Theory and practice*. CAST Professional Publishing.

Mitra, S. (2011, October 13–16). *Responsibility, leadership and education* [Keynote address]. Heads of Schools Conference, Singapore.

Murphy, P. (1999). *Learners, learning & assessment*. Paul Chapman Publishing in association with The Open University.

National Research Council. (2011). *Assessing 21st century skills: Summary of a workshop.* The National Academies Press.

Newton, P. (2011). A level pass rates and the enduring myth of norm-referencing. *Research Matters*. https://doi.org/10.17863/CAM.100441

Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, *14*(2), 149–170. https://doi.org/10.1080/09695940701478321

Newton, P. E. (2012). *We need to talk about validity* [Paper presentation]. The National Council for Measurement in Education Annual Meeting, Vancouver, Canada.

Newton, P. E., & Shaw, S. D. (2014). *Validity in educational and psychological assessment*. London: SAGE Publications.

Peterson, A. D. C. (1971). *New techniques for assessment of pupils' work*. Council of Europe.

Peterson, A. D. C. (2003). *Schools across frontiers: The story of the International Baccalaureate and the United World Colleges* (2nd ed.). Open Court Publishing Company.

Popham, W. J. (1978). *Criterion-referenced measurement*. Prentice Hall.

RAND Corporation. (2012, April). *Teaching and learning 21st century skills: Lessons from the learning sciences*. Asia Society. Retrieved January 1, 2017, from http://www.rand.org/pubs/external_publications/EP51105.html

Rao, K., Currie-Rubin, R., & Logli, C. (2016). *UDL and inclusive practices in IB schools worldwide*. CAST Professional Learning.

Schleicher, A. (2016). *The case for 21st-century learning*. OECD. Retrieved April 4, 2025, from https://web-archive.oecd.org/2012-06-14/61660-thecasefor21st-centurylearning.htm

Shepard, L. A. (1992). Commentary: What policy makers who mandate tests should know about the new psychology of intellectual ability and learning. In B. R. Gifford, & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement and instruction* (pp. 301–328). Kluwer Academic Publishers.

Sireci, S. G. (2007). On validity theory and test validation. *Educational Researcher*, *36*(8), 477–481. https://www.jstor.org/stable/4621103

Smith, D. J., & Tomlinson, S. (1989). *The school effect: A study of multi-racial comprehensives*. Policy Studies Institute.

Snyder, B. R. (1971). *The hidden curriculum*. MIT Press.

Surgenor, P. (2010). *Teaching toolkit: Effect of assessment on learning*. UCD Dublin.

Vygotsky, L. S. (1962). *Thought and language*. MIT Press.

Wiliam, D. (1993). Validity, dependability and reliability in National Curriculum assessment. *The Curriculum Journal, 4*(3), 335–350. https://doi.org/10.1080/0958517930040303

Winkley, J., & Cresswell, M. (2011). Introduction to the concept of reliability. In Ofqual's *Reliability compendium* (Chapter 1). The Office of Qualifications and Examinations Regulation.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Wood, D. (1998). *How children think and learn: The social contexts of cognitive development* (2nd ed.). Blackwell Publishing.

Zanga, G., & De Gioannis, E. (2023). Discrimination in grading: A scoping review of studies on teachers' discrimination in school. *Studies in Educational Evaluation, 78.* https://doi.org/10.1016/j.stueduc.2023.101284

Glossary

The IB is committed to using accessible language and minimizing the use of jargon. To support this commitment, this glossary defines some key words and concepts that appear in this and other assessment-related resources. Some of these terms may also be referenced in communications from the IB.

Term	Definition
Academic integrity	A set of values and skills that promote personal integrity and good practice in learning, teaching and assessment.
Academic misconduct	Behaviour (whether deliberate or inadvertent) that results in, or may result in, the student or any other student gaining an unfair advantage in one or more components of assessment. Behaviour that may disadvantage another student is also regarded as academic misconduct.
Achievement level	The level given when work submitted for assessment reflects the corresponding descriptor. Achievement levels are shown in the left-hand column of assessment criteria.
Achievement level descriptors	Descriptions of the features of work submitted for assessment that are expected at each achievement level.
Admin Console	A secure website used for MYP on-screen examinations. It allows the MYP coordinator and/or designated chief invigilator to complete key administration activities, such as device registration and access to examination packages. The Admin Console is accessed via a link in IBIS. Note: For DPCP digital examinations, see **Digital Examination System**.
Adverse circumstances	Circumstances beyond the control of the student that might be detrimental to their assessment performance. These include severe stress, exceptionally difficult family circumstances, bereavement or events that may threaten the health or safety of students. The same circumstances may affect a group of students, or all students within a school. Adverse circumstances do not include: • shortcomings on the part of the school at which the student is registered • the failure of students to improve performance despite receiving authorized inclusive access arrangements.
Aggregation	The process of combining marks and scores into a final outcome.
Alignment	Agreement in principle and practice between shared values and aspirations for learning (written curriculum), how teachers work (taught curriculum) and what students learn (assessed curriculum).
Analytic markscheme	A markscheme that indicates what the right answer is and where marks should be awarded.
Assessment	Any of the tasks completed by students to demonstrate their ability. These tasks can be school-based and administered, or IB-submitted, and are a collection of evidence upon which judgements can be made on learning and teaching.

Term	Definition
Assessment component	One or more tasks that are collected together, comprising part of the overall assessment, such as an examination, portfolio of work, project or research assignment.
Assessment content developer	A professional who creates, scrutinizes and reviews proposed examination questions and the associated markscheme that will be used for the assessments. Assessment content developers are nearly always senior examiners, but it is a role requiring a discrete set of skills.
Assessment criteria	The criteria against which a student's performance is measured.
Assessment cycle	The steps taken in creating, taking and marking assessments including examinations. It is a cycle because the IB learns from each examination session to improve future sessions.
Assessment response	A term used to describe all material produced by a student in response to an assessment task.
Assessment strategy	The method or approach used when gathering information about student learning (for example, observation, open-ended tasks, selected responses).
Assessment task	The activity or series of activities with which students engage in order for assessment to take place.
Assessment tool	A method of collecting information about a student's performance and understanding.
Attainment-referencing (also **Weak criterion-referencing**)	The comparison of student attainment against pre-defined descriptions of achievement (criteria) and the performance of previous cohorts. This is the approach the IB takes to maintaining standards.
Atypical response	An answer to a task that is significantly different to those usually received. Examples of atypical responses include incomplete work, non-compliant work, unanticipated responses, problematic work or malpractice.
Authentication	The evidence and process of verifying that work has been undertaken by the student. Examples include signatures from the teacher and student confirming that the student's work is their own.
Automarking	The process of using technology to assess students' work against a pre-defined markscheme where an answer is objectively right or wrong.
Backwash effect	The impact that later parts of a process have on the delivery of earlier parts. In an educational context, "backwash" usually refers to the way learning and teaching is affected by assessment.
Bandwidth checker	An online tool used by DPCP school administrators to confirm that the school's internet bandwidth is sufficient to support multiple students sitting DPCP digital examinations at the same time. It is recommended that the bandwidth checker is run well in advance of examination sessions to ensure optimal setup.
Bias	An effect in which a defined group performs differently on a specific question or task than the average for a reason other than ability in the skill or knowledge that is being assessed.
Bring your own device	A practice in which students use their own personal devices for schoolwork. Schools provide specifications to guide parents and legal guardians in purchasing a suitable device for student use in school.
Chief examiner	The most senior examiner who is responsible for ensuring that standards are maintained over time and between disciplines within a subject group.

Term	Definition
Chief invigilator	DPCP: The chief invigilator is responsible for overseeing digital examinations in the school, including completing administrative tasks in the Digital Examination System and managing any technical issues before and during the examination period. They must be a member of the school staff and are assigned this role in My IB by the programme coordinator. This role should be separate from, but work closely with, the programme coordinator to ensure a smooth examination process. In some cases, the programme coordinator may also serve as the chief invigilator. MYP: The role of chief invigilator is optional. The programme coordinator may choose to assign someone to support the completion of administrative tasks related to on-screen examinations.
Command term	The word(s) in a question that explains the assessment objective that is being assessed.
Comparability	The degree to which a particular outcome can be considered the same as another outcome. It is typically used between years or between subjects.
Compatibility checker	A standalone application used in the MYP on-screen and DPCP digital examinations to confirm whether a device meets the technical requirements for running a digital examination. The checker must be run on every device that will be used for assessment. Examinations cannot be taken on devices that do not pass the compatibility check.
Construct relevance	The degree to which the assessment actually tests the skills and knowledge that it is intended to.
Course	A prescribed number of classes, lessons or teaching hours within a defined period of study. Schools organize learning and teaching of subjects through disciplinary and interdisciplinary courses.
Course results	The IB's primary assessment outcome document. It shows each subject the student has taken and the grade achieved (1–7, or A–E where applicable). It also shows other relevant assessment outcomes, such as the grade achieved in the core components. Finally, the document records the student's name, personal code, session number, session in which the awards were achieved, date of issue, name of school registering the student (and replacement, if appropriate).
Criterion-referencing (also **Attainment-referencing**)	A comparison of student attainment against pre-defined descriptions of achievement (criteria) for grading.
Criterion-related assessment	An assessment process based on determining achievement levels against previously agreed criteria. The standard is therefore fixed, rather than variable and depending on the achievement of the entire cohort of students.
Definitive mark	The mark awarded by the principal examiner for a particular piece of work submitted for assessment. This represents the mark that every other examiner should be aiming to replicate. (See also **Quality model**.)
Device	A laptop or desktop computer used for IB digital assessments. The IB uses the term "device" instead of "computer" to reflect ongoing exploration of future compatibility with tablets and other platforms.
Differentiation (assessment)	To distinguish between students demonstrating different levels of competency.

Term	Definition
Differentiation (learning and teaching)	Modifying teaching strategies to meet the needs of diverse learners, through varied content, process and products.
Digital examination responses	In IB digital assessments, a student's completed work is referred to differently depending on the programme. In the MYP, it is known as a "student response file". In the DPCP, it is referred to as a "student response". Both refer to the final digital output produced during the examination and submitted for assessment. In paper-based examinations, this would be called a "script".
Digital Examination System	The online system that schools use to administer DPCP digital examinations.
Digital examinations	In the MYP, on-screen examinations have been in place for over a decade. They are specifically designed for the digital environment and include interactive, media-rich tasks completed in a secure examination setting. In the DPCP, digital examinations are formal, final-year assessments completed by students, using a device under secure examination conditions. They will be phased in, digitally replicating traditional paper examinations.
Discipline	A branch of learning or field of academic study; a way of ordering knowledge for the purpose of instruction (known generally, for practical purposes of assessment in the MYP and DP, as "subjects").
Dynamic sampling	A refinement of the moderation process that allows better use of quality checks.
eAssessment	MYP assessment comprising three components: on-screen assessment, ePortfolio, and personal project.
eMarking	The process by which examiners mark examination material directly on the device screen.
Enquiry upon results	Review of marks undertaken at a school's request.
ePortfolios	The system/process by which schools upload MYP students' internally assessed examination/coursework material to be externally moderated by the IB.
Examination	Examination is one form of assessment and includes a collection of one or more tasks of various types (short answer, extended answer, problem-solving or analytical questions; sometimes practical or oral tasks) that students must respond to under tightly controlled conditions in a set time.
Examination invigilator	An individual who supervises the examination environment and helps to maintain examination security.
Examination paper	The set of tasks and questions that a student is asked to complete in an examination. It may refer to an examination that is taken on a computing device screen, or an examination taken with pen and paper.
Examination session	The period during which examinations are taken and marked. The IB has two examination sessions each year, in May and November.
Examiner	An individual who assigns marks to a student's external assessment.
Examiner re-mark	The process of re-marking an examiner's allocation of student responses (full responses or question item groups, depending on the structure of the assessment) in the case that their marking is found to be inconsistent or

Term	Definition
	deviates significantly from the required standard. This often occurs as a result of moderation failure.
Exceptional circumstances	Circumstances that are not commonly within the experience of other students with assessment access requirements. The IB reserves the right to determine which circumstances qualify as "exceptional" and therefore justify a particular inclusive access arrangement.
External assessment	Assessment that is set and marked by the IB and not by a student's teacher.
External moderation	See **Moderation**.
Externally assessed	Work that is assessed wholly by the IB.
Familiarization	The process by which students learn to navigate the IB's digital examination environment and tools using a content-free practice version of the systems. Schools are responsible for ensuring students have regular access to familiarization tools ahead of their digital examinations. Familiarization environments are available for both MYP and DPCP on the Programme Resource Centre (PRC).
Familiarization tool	A content-free simulation of a digital examination environment. It allows students to practise navigating the platform—exploring tools and understanding the examination layout—without any subject-specific content.
Final assessment	The summative assessment of work at the end of the period of study.
Formative assessment	Ongoing assessment aimed at providing information to guide teaching and improve student performance.
Grade	A description of student achievement. Final grades for work submitted for assessment range from 1 (lowest) to 7 (highest). The grade represents the IB's judgement on the overall qualities that the student has demonstrated, and is consistent between years and subjects. A grade describes how good the student's performance is, and should mean the same for every examination, year and subject. This takes into account the relative difficulty of the component between sessions and any changes in standards.
Grade award	A process through which the matter of how to convert marks into grades is determined. This ensures that grades mean the same thing whichever session a student takes their examination in.
Grade boundary	The point at which student achievement moves from one grade to another. It is often used to indicate the lowest or highest criterion level totals, which corresponds to a particular grade.
Grade descriptors	The articulation of the qualities expected of students to achieve each grade. A grade descriptor may be specific to a subject, specific to a subject group, or general across a whole programme. In each case, a grade descriptor should describe the same characteristics; the more specific examples only explain what these descriptions mean in a subject-specific context.
Holistic criteria	An approach to evaluating a student's work that considers the work as a single outcome, rather than looking at separate elements of it individually (for example, communication, subject knowledge, quality of argument).
IB Information System (IBIS)	A system that allows programme coordinators to complete administrative procedures and obtain news and information from the IB via a password-protected web server.

Term	Definition
Inclusive access	An assessment that has considered the needs of all students, so that students can fairly demonstrate their competence in the subject.
Inclusive access arrangements	Changed or additional conditions during the assessment process for a student with assessment access requirements. These enable the student to demonstrate their level of attainment more fairly.
Interdisciplinary assessment	Combining or involving two or more branches of learning or fields of academic study within a single assessment. In the DP, an interdisciplinary subject is one that meets the requirements of two subject groups through a single subject. In the MYP, interdisciplinary study can be developed both within and between subject groups. MYP external interdisciplinary assessment always involves multiple subject groups.
Internal assessment (IA)	Assessment carried out by teachers in the school.
Internal standardization	The process by which all teachers of a particular subject in a school ensure they are assessing to the same standard.
Internally assessed	Work that is assessed by the students' teachers. Internally assessed material is sampled by the IB for moderation purposes.
Issue of results	The process through which students receive grades from the IB based on their assessments.
Item	The smallest appropriate subdivision of an assessment; a discrete, yet whole, assessment part being tested, as outlined, communicated and expected based on the "Syllabus outline" or "External assessment details" sections of an IB subject guide.
Judgement	The consideration of a student's work against an individual assessment criterion.
Lockdown mode	A secure examination setting that prevents students from accessing other applications, websites or files during a digital examination. MYP: Lockdown functionality is built into the examination package and is activated automatically when the assessment is launched. DPCP: Lockdown mode will be enforced using the Safe Exam Browser for Windows and Mac, and Chrome Kiosk mode for school-managed Chromebooks. Schools are responsible for configuring student devices.
Maladministration	An action by an IB World School that infringes IB rules and regulations, and potentially threatens the integrity of IB examinations and assessments. It can happen before, during or after the completion of the assessment or completion of the examination.
Malpractice	Any practice that subverts the principles of academic integrity (for example, plagiarism).
Manageability	The degree to which an assessment and individual tasks place a burden on the student or school. Examples of manageability include the length of the assessments, the equipment or material required to deliver the assessment or the number of assessments required in a qualification.
Mark(s)	A numerical value that reflects the quality of a student's response to a specific task. Marks are awarded in accordance with criteria, markbands or a markscheme to reflect how much of a question the student has answered correctly. The allocation of marks is different for each question and examination.

Term	Definition
Markbands	A specified range of marks that should be awarded to a student answer that shows certain qualities.
Markscheme	Guidance for awarding criterion levels for a given piece of work.
Missing grade procedure	A mechanism for providing a grade for MYP students where the IB is not able to calculate an accurate or fair grade based on the work the student has completed. This process is appropriate in those circumstances where the lack of evidence is due to the actions of the IB or third parties (not including the school), and where it would not be reasonable for the student to be asked to complete the assessment on another occasion.
Missing mark procedure	A mechanism for providing a mark for DPCP students where the IB is not able to calculate an accurate or fair mark based on the work the student has completed. This process is appropriate in those circumstances where the lack of evidence is due to the actions of the IB or third parties (not including the school), and where it would not be reasonable for the student to be asked to complete the assessment on another occasion.
Moderation	A process to ensure that a common standard of assessment is achieved. This process includes reviewing samples of work submitted for assessment and adjusting assessors' marks where necessary.
Moderation factor	An arithmetical adjustment applied to a teacher's marks (at a criterion level) to bring them in line with the definitive standard.
Moderation sample	The sample of work submitted to the IB to ensure marking is of the required standard.
Modified paper	Changes made to an assessment to allow a student with specific needs to be able to take the assessment on an equal footing with students who do not have these needs. Examples include changing the shape or style of the type font. Such adjustments must not change the nature of the question being asked.
Multiple-choice question	A question in which a student must select the correct answer from a supplied list of possible answers.
MYP on-screen examination	A formal, timed, media-rich digital assessment designed specifically for the MYP. These examinations are completed in a secure digital environment and consist of subject-specific tasks that students respond to on a device. An internet connection is not required to run the examination, but online delivery is strongly recommended. It allows student responses to be saved automatically and reduces administrative workload. Most schools choose to run these assessments online for this reason.
Norm-referencing	The practice by which attainment is determined by comparing (referencing) a student's performance against that of the entire cohort for whom the assessment is designed.
Objective	One of a set of statements describing the skills, knowledge and understanding that will be assessed.
Paper	See **Examination paper**.
Pilot subject	A subject undergoing evaluation. Pending successful evaluation, the pilot subject will become generally available.
Plagiarism	The representation, intentionally or unintentionally, of the ideas, words or work of another person without proper, clear and explicit acknowledgement.

Term	Definition
Practice script	Examples of work submitted for assessment that are identified and marked during standardization; these are given to examiners to demonstrate the standard to which they should mark.
Predictability	The degree to which it is possible to gauge what or when something will happen. In the context of assessment, "predictability" refers to the ability of schools to anticipate questions that will be asked on an examination paper, and when those questions will be asked.
Principal examiner	The person responsible for leading the assessment of a component. The principal examiner sets the standard for the assessment and is usually also one of the assessment authors. In the MYP, the role of principal examiner is slightly different from other examination systems. An MYP principal examiner is the head of a particular discipline and is responsible for leading the team designing the assessment, for setting and maintaining standards and for mentoring examiner team leaders.
Qualification script	An example of work submitted for assessment selected by the principal examiner that is used to check that examiners have understood the required standard of marking before they mark live student scripts.
Quality model	The approach that the IB takes to ensure that students receive the correct assessment outcome. The principal examiner sets the correct standard of response for each question, and each examiner must reproduce this standard. For externally marked assessments, this is done by providing guidance to examiners through standardization, checking their understanding of the standard with qualification scripts, and monitoring their marking regularly through seed scripts.
Question	A task or activity used to allow a student to demonstrate their competence in a subject.
Question bank	A collection of questions, accompanied by information about the relevant topic and expected degree of difficulty. The information in a question bank can be used to create examinations papers. The IB does not currently use question banks.
Question item group	One or more related questions within an examination paper that are considered as a group. Examiners are asked to mark individual question item groups rather than whole examinations. This approach provides more reliable marking than whole-script approaches.
Reliability	The degree to which the student will receive the same outcome every time their work is assessed. It can refer to reliability between examiners or the reliability of a single examiner.
Response language	The language in which a student answers an assessment.
Retake	A second or subsequent attempt at one or more examinations with the aim of obtaining an MYP Certificate or IB Diploma, or of increasing the total mark on a certificate already received.
Script	A student's responses to a paper-based examination. It can also be any work that has been submitted for assessment.
Seed	A script that has been marked by the principal examiner and is randomly added to a batch of scripts allocated to an examiner for marking. The marks the examiner awards the seed script are checked against those given by the principal examiner to ensure the examiner is marking to the set standard.

Term	Definition
	Dynamic sampling moderation seeds are used in the same way as part of the moderation process.
Senior examiner	An experienced examiner who supports the principal examiner.
Session	See **Examination session**.
Specimen examinations	Full, subject-based examples of IB examinations. Specimen examinations contain real assessment content and mirror the format and level of the final assessments. They are typically used to help students practise under examination-like conditions. MYP: Specimen packages are available on the PRC. DPCP: Specimens will be made available through the PRC. For digital examinations, additional access may be provided via the live Digital Examination System.
Standard	The expected level of performance that corresponds to a particular score, grade or outcome.
Standardization	The collaborative process by which a common standard of assessment is achieved among moderators or examiners.
Standardization meeting	A meeting held by the principal examiners to describe the required standard for marking and set seed scripts.
Student	A person who is taking part in an IB course or educational programme.
Student registration	The process undertaken by a programme coordinator to register students for IB assessment.
Student response file	In MYP on-screen examinations, a "student response" refers to the student's completed work—which may include multimedia—produced during the examination. The process of creating the response is part of what is being assessed. The student response is submitted as an electronic file, often referred to as the "student response file".
Submission	The process by which a student (or school on behalf of the student) hands in their final work to the IB.
Summative assessment	Assessment aimed at determining a student's competency or achievement level generally at the end of a course of study, or at the end of a unit of work.
System requirements	The minimum technical specifications a device must meet to run IB digital examinations smoothly and securely. These requirements are in place to ensure that students are not disadvantaged by hardware issues during their assessments. Requirements are published for both the MYP and the DPCP. Schools are responsible for ensuring that all student devices meet the necessary specifications.
System requirements checker	A tool used to verify that a device meets the technical requirements for running a DPCP digital examination. It helps schools test devices in advance and confirm that they will support a smooth and secure assessment experience. This checker is used during examination preparation, not on examination day, and complements the published system requirements.
Teacher support material	Additional information to help teachers understand what is required by an IB course. Teacher support material is intended to give practical advice and to aid

Term	Definition
	understanding and implementation of the theory presented in the subject guides.
Team leader	An examiner who leads a team of examiners.
Test	"Test" can be used in a generic sense to refer to an examination, although it is sometimes given a specific meaning in academic papers on assessment. Sometimes the overall assessment of a student will be broken down into several separate pieces taken at different times. These are called "assessment components" or "components". Examples of components might be oral examinations, any individual examination or an IA.
Tolerance	The small variation from the principal examiner's definitive mark that the IB believes is close enough to show the examiner is still marking to the correct standard. Tolerances are necessary because marking is a matter of judgement; even experienced examiners will vary slightly when re-marking the same piece of work. Tolerances vary according to the number of marks, the kind of question and the subject.
Universal design for assessment	The concept that all assessments should be developed with an understanding of the range of requirements that students may have, rather than treating some students differently. This is part of the IB's commitment to universal design for learning.
Universal design for learning	An orientation and framework for student learning that focuses on all students. It aims to create inclusive environments by providing multiple means of engagement, representation and expression.
Validity	The term that describes whether an assessment, or the purpose for which the assessment results are being used, is fit for purpose.
Validity argument	The evidence and explanation for decisions made in creating an assessment that justify that it is fit for purpose.
Weak criterion-referencing (also **Attainment-referencing**)	The comparison of student attainment against pre-defined descriptions of achievement (criteria) and the performance of previous cohorts. This is the approach the IB takes to maintaining standards.
Working languages	The languages in which the IB communicates with its stakeholders and in which it is committed to providing a range of services for the implementation of its programmes. The IB's current working languages are English, French and Spanish.

Printable resources

Throughout this publication, there are references to printable resources. All printable resources are listed below for ease of access.

- "Accountabilities and responsibilities of the key roles in the assessment cycle" (PDF)
- "Bloom's taxonomy" (PDF)
- "Building an ethical mindset" (PDF)
- "Implementing inclusive access arrangements" (PDF)
- "Marking internal assessment: Expectations of teachers in the moderation process" (PDF)
- "Moderation using dynamic sampling" (PDF)
- "Predicted grades: A teacher's guide" (PDF)
- "Range of quality model IA scripts required" (PDF)
- "Summative and formative assessment" (PDF)
- "The IB's principles of assessment" (PDF)
- "The validity chain" (PDF)