



International Baccalaureate®
Baccalauréat International
Bachillerato Internacional

Principes et pratiques de l'évaluation – Des évaluations de qualité à l'ère du numérique



International Baccalaureate®
Baccalauréat International
Bachillerato Internacional

Principes et pratiques de l'évaluation – Des évaluations de qualité à l'ère du numérique

Principes et pratiques de l'évaluation – Des évaluations de qualité à l'ère du numérique

Version française de la publication parue originalement en anglais en juillet 2019 sous le titre *Assessment principles and practices—Quality assessments in a digital age*

Publiée en juillet 2019

Mise à jour en novembre 2021

Publiée pour le compte de l'Organisation du Baccalauréat International, fondation éducative à but non lucratif sise 15 Route des Morillons, CH-1218 Le Grand-Saconnex, Genève, Suisse, par

International Baccalaureate Organization (UK) Ltd
Peterson House, Malthouse Avenue, Cardiff Gate
Cardiff, Pays de Galles CF23 8GL
Royaume-Uni
Site Web : <https://ibo.org/fr/>

© Organisation du Baccalauréat International 2019

L'Organisation du Baccalauréat International (couramment appelée l'IB) propose quatre programmes d'éducation stimulants et de grande qualité à une communauté mondiale d'établissements scolaires, dans le but de bâtir un monde meilleur et plus paisible. Cette publication fait partie du matériel publié pour appuyer la mise en œuvre de ces programmes.

L'IB peut être amené à utiliser des sources variées dans ses travaux, mais vérifie toujours l'exactitude et l'authenticité des informations employées, en particulier dans le cas de sources participatives telles que Wikipédia. L'IB respecte les principes de la propriété intellectuelle et s'efforce toujours d'identifier les détenteurs des droits relatifs à tout matériel protégé par le droit d'auteur et d'obtenir d'eux, avant publication, l'autorisation de réutiliser ce matériel. L'IB tient à remercier les détenteurs de droits d'auteur qui ont autorisé la réutilisation du matériel apparaissant dans cette publication et s'engage à rectifier dans les meilleurs délais toute erreur ou omission.

Le générique masculin est utilisé ici sans aucune discrimination et uniquement pour alléger le texte.

Dans le respect de l'esprit international cher à l'IB, le français utilisé dans le présent document se veut standard et compréhensible par tous, et non propre à une région particulière du monde.

Tous droits réservés. Aucune partie de cette publication ne peut être reproduite, mise en mémoire dans un système de recherche documentaire, ni transmise sous quelque forme ou par quelque procédé que ce soit, sans autorisation écrite préalable de l'IB ou sans que cela ne soit expressément autorisé par le [règlement de l'IB en matière d'utilisation de sa propriété intellectuelle](#).

Vous pouvez vous procurer les articles et les publications de l'IB sur le [magasin en ligne de l'IB](#) (adresse électronique : sales@ibo.org). Toute utilisation commerciale des publications de l'IB (qu'elles soient commerciales ou comprises dans les droits et frais) par des tiers exerçant dans le milieu de l'IB mais sans relation formelle avec lui (ce qui comprend notamment les organisations spécialisées dans le tutorat, les fournisseurs de perfectionnement professionnel, les éditeurs dans le domaine de l'éducation et les acteurs chargés de la planification de programmes d'études ou de la gestion de plateformes numériques contenant des ressources à l'intention des enseignants) est interdite et nécessite par conséquent l'obtention d'une licence écrite accordée par l'IB. Veuillez envoyer toute demande de licence à l'adresse copyright@ibo.org. Des informations complémentaires sont disponibles sur le site Web public de l'IB.

Déclaration de mission de l'IB

Le Baccalauréat International (IB) a pour but de développer chez les jeunes la curiosité intellectuelle, les connaissances et la sensibilité nécessaires pour contribuer à bâtir un monde meilleur et plus paisible, dans un esprit d'entente mutuelle et de respect interculturel.

À cette fin, l'IB collabore avec des établissements scolaires, des gouvernements et des organisations internationales pour mettre au point des programmes d'éducation internationale stimulants et des méthodes d'évaluation rigoureuses.

Ces programmes encouragent les élèves de tout pays à apprendre activement tout au long de leur vie, à être empreints de compassion, et à comprendre que les autres, en étant différents, puissent aussi être dans le vrai.



Profil de l'apprenant de l'IB

Tous les programmes de l'IB ont pour but de former des personnes sensibles à la réalité internationale, conscientes des liens qui unissent entre eux les humains, soucieuses de la responsabilité de chacun envers la planète et désireuses de contribuer à l'édification d'un monde meilleur et plus paisible.

En tant qu'apprenants de l'IB, nous nous efforçons d'être :

CHERCHEURS

Nous cultivons notre curiosité tout en développant des capacités d'investigation et de recherche. Nous savons apprendre indépendamment et en groupe. Nous apprenons avec enthousiasme et nous conservons notre plaisir d'apprendre tout au long de notre vie.

INFORMÉS

Nous développons et utilisons une compréhension conceptuelle, en explorant la connaissance dans un ensemble de disciplines. Nous nous penchons sur des questions et des idées qui ont de l'importance à l'échelle locale et mondiale.

SENSÉS

Nous utilisons nos capacités de réflexion critique et créative, afin d'analyser des problèmes complexes et d'entreprendre des actions responsables. Nous prenons des décisions réfléchies et éthiques de notre propre initiative.

COMMUNICATIFS

Nous nous exprimons avec assurance et créativité dans plus d'une langue ou d'un langage et de différentes façons. Nous écoutons également les points de vue d'autres individus et groupes, ce qui nous permet de collaborer efficacement avec eux.

INTÈGRES

Nous adhérons à des principes d'intégrité et d'honnêteté, et possédons un sens profond de l'équité, de la justice et du respect de la dignité et des droits de chacun, partout dans le monde. Nous sommes responsables de nos actes et de leurs conséquences.

OUVERTS D'ESPRIT

Nous portons un regard critique sur nos propres cultures et expériences personnelles, ainsi que sur les valeurs et traditions d'autrui. Nous recherchons et évaluons un éventail de points de vue et nous sommes disposés à en tirer des enrichissements.

ALTRUISTES

Nous faisons preuve d'empathie, de compassion et de respect. Nous accordons une grande importance à l'entraide et nous œuvrons concrètement à l'amélioration de l'existence d'autrui et du monde qui nous entoure.

AUDACIEUX

Nous abordons les incertitudes avec discernement et détermination. Nous travaillons de façon autonome et coopérative pour explorer de nouvelles idées et des stratégies innovantes. Nous sommes ingénieux et nous savons nous adapter aux défis et aux changements.

ÉQUILIBRÉS

Nous accordons une importance équivalente aux différents aspects de nos vies – intellectuel, physique et affectif – dans l'atteinte de notre bien-être personnel et de celui des autres. Nous reconnaissons notre interdépendance avec les autres et le monde dans lequel nous vivons.

RÉFLÉCHIS

Nous abordons de manière réfléchie le monde qui nous entoure, ainsi que nos propres idées et expériences. Nous nous efforçons de comprendre nos forces et nos faiblesses afin d'améliorer notre apprentissage et notre développement personnel.

Le profil de l'apprenant de l'IB incarne dix qualités mises en avant par les écoles du monde de l'IB. Nous sommes convaincus que ces qualités, et d'autres qui leur sont liées, peuvent aider les individus à devenir des membres responsables au sein des communautés locales, nationales et mondiales.

Table des matières

Introduction et présentation	1
Introduction	1
Quels programmes de l'IB sont concernés ?	4
Pourquoi lire ce document ?	5
À qui s'adresse cette ressource ?	6
Comment utiliser cette ressource : plusieurs pistes possibles	7
Quel est le sens de la formule « Des évaluations de qualité à l'ère du numérique » ?	13
Quel est le lien entre cette ressource et les autres publications de l'IB ?	14
Une évaluation s'appuyant sur la technologie	17
En quoi consistent les principes et les pratiques de l'évaluation ?	21
Terminologie de l'évaluation	22
Les nouveaux termes pour l'évaluation électronique	24
Section A – Les principes de l'évaluation	26
En quoi consiste l'évaluation ?	26
Quelle est l'adéquation avec l'objectif visé ? Comprendre la validité	32
Éléments de la chaîne de validité	38
Définition des normes	56
Description de la réussite : les résultats des candidats à l'évaluation sommative	61
Notation des évaluations	65
Qu'est-ce qu'une bonne évaluation ?	71
Les principes de l'IB en matière d'évaluation	84
Section B – Les pratiques de l'évaluation de l'IB	85
Qu'entend-on par « pratique » ?	85
Transmission des résultats des candidats	86
Processus d'évaluation – Rôles et responsabilités	95
Intégrité de l'évaluation	100
L'équité pour tous : répondre aux besoins des candidats	104
Le cycle d'évaluation	109
Préparation des épreuves d'examen : élaboration et qualité	112
Examens	117
Établissement de la norme : comment préparer les examinateurs à la notation	121
Notation	126
Révision de notation	139
L'attribution des notes finales (et la totalisation)	146
Les contrôles de la qualité fondés sur les copies dites « à risque »	159
Le comité d'attribution des notes finales	160

La préparation de la publication des résultats	162
L'équité pour tous : répondre aux besoins des candidats	167
Les réclamations concernant les résultats, la procédure d'appel et le retour d'information général	172
La préparation des évaluations de l'année suivante	177
Les commentaires destinés aux établissements	180
Section C – Les processus spécifiques aux programmes de l'IB	182
<hr/>	
En quoi consistent les processus spécifiques aux programmes ?	182
Le socle commun à tous les programmes	184
Les compétences de l'élève et le profil de l'apprenant	186
Le Programme du diplôme de l'IB	192
Le Programme à orientation professionnelle de l'IB	199
Le Programme d'éducation intermédiaire de l'IB	203
Le Programme primaire de l'IB	208
Annexes	210
<hr/>	
Annexe 1 – Révision de notation de l'évaluation interne	210
Annexe 2 – Feuille de route pour créer un argument de validité	216
Bibliographie	219
Glossaire	222

Introduction

Le présent document a pour but d'expliquer les principes auxquels le Baccalauréat International (IB) a adhéré pour s'assurer qu'il procède à une évaluation véritablement utile, juste et privilégiant l'intérêt supérieur des élèves concernés.

Les enseignants sont les premiers destinataires visés, mais le contenu proposé devrait être à la portée des élèves, des examinateurs, des parties prenantes et de toute autre partie intéressée.

Le but de l'évaluation électronique est de pouvoir évaluer ce qui importe vraiment (et pas simplement ce qui peut être évalué avec du papier et un crayon), et ce, en utilisant le support qui est le plus familier aux élèves d'aujourd'hui.

Ce qui compte ne peut pas toujours être compté, et ce qui peut être compté ne compte pas forcément.

(Einstein ou Cameron, 1963)

À l'évidence, si les autres critères sont moins fiables que les examens, compter davantage sur ces critères signifiera faire reposer la sélection sur des décisions moins fiables.

(Cresswell, 1986, p. 37 – 54)

Ces deux citations illustrent l'ampleur du défi que présente l'évaluation. Nombre des objectifs du système éducatif du Baccalauréat International (IB) sont difficiles à évaluer, mais si les apprenants ne sont pas évalués avec précision, les décisions importantes qui se répercuteront sur leur vie reposeront sur une base moins juste et moins fiable.

Les programmes de l'IB sont enseignés dans plus de 140 pays par des établissements scolaires qui présentent un large éventail de contextes et de traditions pédagogiques. La philosophie et les approches adoptées par l'IB pour évaluer les élèves pourront sembler familières à certains établissements, alors que d'autres pourront voir des zones d'ombre dans ce système.

Il paraît tout particulièrement utile d'apporter ces éclaircissements en période de changement et compte tenu du fait que les effets de la technologie sur l'éducation, et notamment sur l'évaluation, pourront encore être perçus au cours des dix prochaines années. L'IB est intimement convaincu que la technologie doit être mise au service de l'évaluation, et l'évolution vers des évaluations sur ordinateur n'altérera en rien ses principes. Elle pourra cependant lui donner de nouvelles occasions de convertir ces principes en applications concrètes. Vous trouverez de plus amples informations à ce sujet dans la section « [Une évaluation s'appuyant sur la technologie](#) ».

Il est essentiel que chacun au sein de la communauté de l'IB comprenne comment fonctionne notre processus d'évaluation externe, quelles sont ses forces et ses limites, et les raisons qui justifient les décisions prises. Une plus grande transparence ne peut que conduire à une meilleure compréhension et, en fin de compte, à une meilleure éducation pour les élèves. Grâce aux possibilités offertes par les ressources informatisées, l'IB espère mettre à la disposition des enseignants un guide clair et accessible, qui présente les informations détaillées dont ils ont besoin pour comprendre les évaluations de l'IB.

Figure 1

Toutes les évaluations nécessitent de trouver un juste équilibre.



Toutes les évaluations nécessitent de trouver un juste équilibre entre des exigences contradictoires, et un grand nombre d'inquiétudes émises au sujet des tests négligent cet aspect.

Il existe, par exemple, une tension entre le souhait de réduire la charge associée aux évaluations et le risque pour les candidats de n'avoir qu'une occasion de montrer leurs aptitudes.

Pour de plus amples informations, veuillez vous reporter à la section « [Quelle est l'adéquation avec l'objectif visé ? Comprendre la validité](#) ».

Un juste équilibre à trouver entre des priorités contradictoires

La notion d'équilibre englobe un autre aspect. Pour l'IB, le développement des élèves passe par un programme d'études global, ce qui doit être pris en compte dans l'approche de l'évaluation. Chaque décision doit donc être prise en tenant compte des incidences sur le programme dans sa globalité, et non en s'intéressant exclusivement à une matière, une discipline ou une évaluation.

Pour bien comprendre les principes de l'IB en matière d'évaluation des élèves et le mode de fonctionnement de cette évaluation, une courte présentation de l'évolution des pratiques de l'évaluation, au gré du temps et des théories émergentes, s'impose. Même s'il est impossible de traiter en profondeur ici tous les points cruciaux, il est important de les mettre en valeur, du fait de leur grande influence sur les

pratiques actuelles. Si vous souhaitez en savoir plus à ce sujet, vous pourrez compléter vos connaissances en lisant les articles de recherche cités dans le texte.

Il a déjà été indiqué que l'IB s'était fixé une tâche délicate, à savoir viser ce qu'il est important d'évaluer et non ce qu'il est facile d'évaluer. Cette idée est probablement exprimée avec plus d'éloquence par Alec Peterson, le premier directeur général de l'IB :

Nous avons besoin d'un processus d'évaluation qui soit aussi valide que possible, dans le sens où il évalue réellement toute la capacité et la personnalité de l'élève par rapport à la prochaine étape de sa vie, mais qui soit en même temps suffisamment fiable pour que les élèves, les parents, les enseignants et les établissements d'accueil soient assurés que justice est faite. De plus, un tel processus ne doit pas, en raison de son effet de remous, fausser la qualité de l'enseignement, être trop lent ou absorber une trop grande quantité de nos ressources scolaires, qui sont peu abondantes.

(Peterson, 1971, p. 27 – 55)

Il est espéré que cette ressource parviendra, au fil des pages, à expliquer comment l'IB pense pouvoir réaliser cet objectif ambitieux et servir ses intentions plus larges en matière d'éducation pour offrir à ses élèves une expérience parmi les meilleures au monde. Si certaines de vos questions ne trouvent pas de réponse dans la présente ressource, n'hésitez pas à contacter le personnel de l'IB chargé de l'évaluation par courriel à l'adresse électronique assessment@ibo.org.

Quels programmes de l'IB sont concernés ?

Par le passé, les documents consacrés aux principes et aux pratiques de l'évaluation concernaient le Programme du diplôme. Cependant, l'éventail de programmes proposés par l'IB s'étant élargi, il semble maintenant opportun d'expliquer en détail la philosophie de l'IB en matière d'évaluation qui est adoptée dans chaque programme : le Programme du diplôme, le Programme à orientation professionnelle (POP), le Programme d'éducation intermédiaire (PEI) et le Programme primaire (PP). Les raisons de l'adoption des principes de l'évaluation de l'IB seront examinées en profondeur, mais la mise en œuvre de ces principes dans chacun des programmes ne sera pas abordée dans le détail. Pour de plus amples informations à ce sujet, vous devrez vous référer au document intitulé *Des principes à la pratique*, publié pour le PEI, le Programme du diplôme et le POP.

Le présent document se concentre plus particulièrement sur l'évaluation sommative externe de l'IB. À l'heure actuelle, l'IB procède en effet à une évaluation sommative externe pour le Programme du diplôme, le POP et le PEI.

Bien que l'évaluation constitue un élément important du PP, ce programme ne se prête pas à une évaluation sommative externe de l'IB. Pour une meilleure compréhension de la philosophie du PP et une description plus détaillée de l'évaluation dans ce programme, veuillez vous référer à la section consacrée à [l'évaluation dans le cadre du PP](#) sur le Centre de ressources pédagogiques.

Les principes de l'évaluation de l'IB abordent une multitude d'objectifs d'évaluation, notamment l'évaluation formative, qui conduisent aux pratiques choisies pour les composantes évaluées par l'IB. Ces principes présenteront un intérêt pour tous les programmes.

Pourquoi lire ce document ?

Parce que les résultats de l'évaluation ont un impact sur la vie des élèves !

La majeure partie du présent document traite de la manière dont l'IB évalue les candidats pour aboutir aux résultats du Programme du diplôme et du Programme à orientation professionnelle (POP), et de façon facultative, aux résultats du Programme d'éducation intermédiaire (PEI). Les élèves s'appuient ensuite sur ces résultats pour poursuivre leurs études ou rejoindre le monde professionnel.

Comme tout système, l'évaluation n'est qu'un outil qui peut conduire à des résultats positifs ou négatifs pour les personnes qui s'y soumettent. Que vous soyez enseignant(e), parent ou élève, vous êtes concerné(e) par les évaluations et devez comprendre de ce fait leurs points forts et leurs points faibles ainsi que les décisions qui doivent être prises par leurs concepteurs et utilisateurs. Vous trouverez ci-après quelques exemples de questions et de remarques courantes envoyées par les enseignants, et des explications quant à la manière dont le présent document y répond.

« Pourquoi l'IB intervient-il ainsi ? »

Les enseignants et les élèves s'interrogent régulièrement sur les motivations à l'origine de l'évaluation externe. À la lecture de ce document, vous devriez comprendre les principes fondateurs des pratiques de l'évaluation de l'IB.

« Ce n'est pas juste, Monsieur. J'avais besoin d'une meilleure note finale ! »

Les résultats aux examens peuvent conditionner beaucoup de choses pour les élèves (par exemple, l'entrée à l'université et leur future carrière), et ces derniers ont besoin de comprendre les motifs des décisions. L'évaluation est une question d'équilibre entre des exigences contradictoires, et le présent document vous aidera à expliquer aux élèves les incidences plus larges d'un contournement des règles en leur faveur, voire à les convaincre que leur note finale est « juste ».

« Je n'avais jamais raisonné de cette manière avant. »

Même pour les enseignants qui exercent depuis plusieurs années, l'évaluation externe peut souvent apparaître comme un processus mystérieux et opaque. Personne ne devrait « subir » aucun aspect d'un système éducatif, et encore moins de celui de l'IB. Il convient donc de pouvoir expliquer aux élèves comment et pourquoi ils se voient attribuer des résultats en particulier.

« Je veux simplement me servir des évaluations pour mieux décider de mon enseignement. »

Même si vous souhaitez utiliser l'évaluation dans un but formatif, il est important que vous compreniez ce qui fait une bonne évaluation et les décisions que cela entraîne. Des évaluations de mauvaise qualité conduiront à des résultats de mauvaise qualité, et cela est également vrai si l'idée est de soutenir l'apprentissage ou tout autre objectif interne à l'établissement, comme une sélection ou une reconnaissance officielle. Le présent document vous aidera à comprendre comment l'IB atteint son but, qui est de proposer une évaluation de qualité venant soutenir ses objectifs pédagogiques.

À qui s'adresse cette ressource ?

L'idée de départ est que le lecteur est un enseignant accoutumé à l'IB et à ses évaluations, qui n'a jamais étudié la théorie de l'évaluation en bonne et due forme. Il semble que ce postulat ouvre la présente ressource au lectorat le plus large possible.

Élèves, parents, chefs d'établissement et coordonnateurs de l'IB pourront y puiser des informations utiles, d'autant plus qu'y figurent un [glossaire](#) et des liens vers d'autres documents permettant une parfaite compréhension de tous les processus engagés.

De même, les sections consacrées aux principes qui sous-tendent les pratiques de l'évaluation de l'IB seront utiles pour comprendre le choix de l'approche adoptée. Les sections de présentation auront un intérêt particulier pour les examinateurs qui pourront replacer leur rôle dans le contexte plus large des évaluations de l'IB.

Enfin, les agents au service des admissions universitaires et les autres interlocuteurs pourront comprendre ce vers quoi tend l'IB au travers de ses évaluations et ce que représentent les résultats des élèves.

Comment utiliser cette ressource : plusieurs pistes possibles

- Dans le présent document, les thèmes sont organisés de manière à amener facilement les lecteurs vers les sections présentant le plus d'intérêt pour eux.
- Le lecteur pourra se servir des **questions thématiques** ci-après pour en apprendre davantage sur les aspects qui comptent le plus pour lui.

Cette ressource a été conçue pour proposer de multiples pistes qui serviront les intérêts et répondront aux besoins des différents lecteurs. La table des matières ci-après est fournie uniquement à titre de référence et vise à dresser la liste de toutes les sections plutôt qu'à suggérer l'ordre dans lequel il faudrait les lire.

En quoi consistent ces questions thématiques ?

Le présent document a pour objectif de donner un aperçu complet de la manière dont l'IB aborde l'évaluation. De nombreux enseignants ne rechercheront pas cette amplitude, mais des informations précises sur certaines parties du processus. À cette fin, plusieurs listes de sections ont été préparées pour couvrir les questions particulières que les enseignants peuvent se poser.

Qu'est-ce que l'évaluation sur ordinateur va changer dans l'évaluation de l'IB ?

Public visé : les enseignants connaissant bien le processus d'évaluation de l'IB, mais désireux de savoir comment l'IB gèrera la transition vers l'évaluation sur ordinateur.

Exemples de questions courantes :

- En quoi consiste l'évaluation sur ordinateur ?
- La notation se fera-t-elle par ordinateur ?
- En quoi cela constitue-t-il une amélioration ?
- Les normes seront-elles les mêmes ?

Vous pourrez consulter une liste de sections susceptibles de présenter un intérêt particulier pour vous en cliquant ici.

Une évaluation s'appuyant sur la technologie	À quoi ressemble l'évaluation sur ordinateur ?	Les risques à éviter avec l'évaluation sur ordinateur
Terminologie de l'évaluation	Quels sont les principaux termes à comprendre ?	Candidat et élève
Évaluation, examen, test et composante	Notation et attribution d'une note finale	Question et élément
Auteur d'épreuve et examinateur	Les nouveaux termes pour l'évaluation électronique	Évaluation sur ordinateur
Fichier de réponses et réponse du candidat	Outil de familiarisation	En cas de doute sur le sens d'un terme
Qu'est-ce qu'une bonne évaluation ?	À quoi ressemble une bonne évaluation sur ordinateur ?	Le cycle d'évaluation

L'impact de l'évaluation électronique sur le cycle d'évaluation	L'évolution vers une évaluation sur ordinateur	Les évaluations sur ordinateur
---	--	--------------------------------

Comment les évaluations sont-elles notées et comment les notes finales sont-elles attribuées ?

Les enseignants trouveront ici des explications sur ce qui se passe une fois qu'ils ont envoyé les travaux des candidats à l'IB pour qu'ils soient notés et sur le fonctionnement du système d'attribution des notes finales.

Exemples de questions courantes :

- Pourquoi les seuils d'attribution des notes finales sont-ils différents cette année ?
- Pourquoi ne puis-je pas avoir un autre examinateur ?
- Qui note les copies ?
- Comment la notation des examinateurs est-elle vérifiée ?
- Comment puis-je déposer un recours ?
- J'aurais attribué le point au candidat. En quoi cela est-il juste ?
- Le barème de notation indique que la réponse du candidat vaut un point. En quoi cela peut-il être juste ?
- La meilleure solution pour les élèves ne serait-elle pas d'avoir une « salle de classe sans note finale » ?

Vous pourrez consulter une liste de sections susceptibles de présenter un intérêt particulier pour vous en cliquant ici.

Terminologie de l'évaluation	Quels sont les principaux termes à comprendre ?	Candidat et élève
Évaluation, examen, test et composante	Notation et attribution d'une note finale	Question et élément
Transmission des résultats des candidats	Qu'entend-on par « note finale de l'IB » ?	Quelle est la différence entre les notes et les notes finales ?
La tyrannie des notes finales : un moindre mal	Qu'est-ce qu'une session d'examens réussie ?	L'accomplissement ne se réduit pas aux notes finales
Processus d'évaluation – Rôles et responsabilités	L'examineur principal et l'examineur en chef	Les autres examinateurs
Les responsabilités du personnel de l'IB	La hiérarchie des examinateurs	Éléments de la chaîne de validité
Établissement de la norme : comment préparer les examinateurs à la notation	La réunion de normalisation	Les copies d'entraînement
Les copies de validation	Les copies de contrôle	La marge de tolérance
La réussite de l'établissement de la norme	Notation	Comprendre la notation comme le résultat du jugement cohérent des examinateurs
Les « notes définitives »	Notes et notes finales sont deux choses différentes	La notation sur ordinateur (et ses principes fondamentaux)
Les différents types de barèmes de notation	Les barèmes de notation analytiques	Les critères globaux – Les bandes de notation

Un soutien supplémentaire pour les examinateurs	Les groupes d'éléments de question (QIG pour Question Item Group en anglais)	Modèle d'assurance de la qualité
Les copies d'entraînement	Les copies de validation	Les copies de contrôle
La marge de tolérance	Les copies difficiles ou inhabituelles	Liens avec un établissement scolaire
La totalisation	L'attribution des notes finales (et la totalisation)	En quoi consiste l'attribution des notes finales ?
Les seuils d'attribution des notes finales fondés sur une appréciation et calculés par interpolation	L'impact de l'évaluation électronique sur l'attribution des notes finales	Les preuves utilisées pour l'attribution des notes finales
L'examen de la cohorte de l'année en cours	Le retour d'information sur l'évaluation	La recherche de preuves dans les copies
L'analyse des statistiques sur les résultats	Trouver un juste équilibre entre toutes les preuves	Les descripteurs des notes finales
Les seuils d'attribution des notes finales fixes	La totalisation	Les contrôles de la qualité de l'attribution des notes finales et les rapports sur la répartition
L'octroi du diplôme ou du certificat du programme	Les observateurs	Les principes de l'attribution des notes finales
Les contrôles de la qualité fondés sur les copies dites « à risque »		

Quelle est l'approche de l'évaluation adoptée par l'IB ?

Public visé : le personnel chargé des admissions universitaires, les enseignants et les interlocuteurs qui sont intéressés par les bases plus théoriques des évaluations de l'IB, la définition donnée des évaluations de qualité, ce qui guide les décisions en matière d'établissement des processus et la manière de trouver un juste équilibre entre les exigences contradictoires de l'évaluation.

Exemples de questions courantes :

- En quoi consistent exactement les évaluations de l'IB et en quoi sont-elles si particulières ?

Vous pourrez consulter une liste de sections susceptibles de présenter un intérêt particulier pour vous en cliquant ici.

En quoi consiste l'évaluation ?	Qu'est-ce que « l'évaluation » ?	Évaluation formative, sommative ou en tant que forme d'apprentissage... Pourquoi évaluer ?
Effet de remous et apprentissage	Notes et notes finales sont deux choses différentes	Quelle est l'adéquation avec l'objectif visé ? Comprendre la validité
Qu'est-ce que la validité ?	Qu'est-ce qui est valide : l'évaluation ou son utilisation ?	Création d'un argument de validité
Maintien de la validité	En quoi l'ordinateur peut-il renforcer la validité ? Quels avantages l'évaluation	Éléments de la chaîne de validité

	électronique présente-t-elle en matière de validité ?	
Trouver un équilibre entre les différents aspects de la validité	Fiabilité	Résultats cohérents et bons résultats sont deux choses différentes
Pertinence de construit et authenticité	Capacité de gestion	Équité et biais
Comparabilité	Approche de la validité adoptée par l'IB	Définition des normes
Les trois sens donnés aux normes	L'approche normative et l'approche critériée des normes de résultats	En quoi consiste l'approche normative ?
En quoi consiste l'approche critériée ?	Quelle est l'approche adoptée par l'IB ?	Le maintien des normes
Description de la réussite : les résultats des candidats à l'évaluation sommative	La tyrannie des notes finales : un moindre mal	L'importance du jugement professionnel
Notation des évaluations	Qu'entend-on par « notation » ?	Les autres formes de notation
La notation et l'évaluation formative	Qu'est-ce qu'une bonne évaluation ?	Une bonne évaluation soutient les objectifs pédagogiques
Qu'est-ce qu'une bonne prévisibilité ?	Une bonne évaluation s'appuie sur tout un éventail de tâches d'évaluation	Le rôle de l'évaluation se déroulant en classe et de l'évaluation interne
Le travail en collaboration et les notes individuelles	Une bonne évaluation tient compte des compétences plus larges du candidat et des compétences de pensée de haut niveau	Les compétences cognitives de haut niveau
Les compétences de l'élève et le profil de l'apprenant	La sensibilité internationale et la compréhension interculturelle	À quoi ressemble une bonne évaluation sur ordinateur ?
Les principes de l'IB en matière d'évaluation	Transmission des résultats des candidats	Qu'entend-on par « note finale de l'IB » ?
Quelle est la différence entre les notes et les notes finales ?	La tyrannie des notes finales : un moindre mal	Qu'est-ce qu'une session d'examens réussie ?
L'accomplissement ne se réduit pas aux notes finales		

Comment sont conçues et rédigées les épreuves d'examen de l'IB ?

Public visé : les enseignants désireux de comprendre le mécanisme d'élaboration des épreuves d'examen ou cherchant à rassurer des personnes ayant des inquiétudes quant à leur qualité.

Exemples de questions courantes :

- Qui rédige les épreuves d'examen ?
- Comment le travail de ces personnes est-il vérifié ?
- Les épreuves d'examen préparées dans différentes langues sont-elles identiques ?
- Testent-elles réellement ce qu'il faut ?

Vous pourrez consulter une liste de sections susceptibles de présenter un intérêt particulier pour vous en cliquant ici.

En quoi consiste l'évaluation ?	Qu'est-ce que « l'évaluation » ?	Évaluation formative, sommative ou en tant que forme d'apprentissage... Pourquoi évaluer ?
Effet de remous et apprentissage	Notes et notes finales sont deux choses différentes	Quelle est l'adéquation avec l'objectif visé ? Comprendre la validité
Qu'est-ce que la validité ?	Qu'est-ce qui est valide : l'évaluation ou son utilisation ?	Qu'est-ce qu'une bonne évaluation ?
Une bonne évaluation soutient les objectifs pédagogiques	Qu'est-ce qu'une bonne prévisibilité ?	Une bonne évaluation s'appuie sur tout un éventail de tâches d'évaluation
Le rôle de l'évaluation se déroulant en classe et de l'évaluation interne	Les rôles se rapportant à la rédaction des épreuves d'examen	Préparation des épreuves d'examen : élaboration et qualité
Les règles qui régissent la rédaction des épreuves d'examen	Les listes de questions	La présentation de la préparation des épreuves d'examen
Le processus conduisant à la validation des ressources	Le processus conduisant à la validation du contenu	Le processus conduisant à la validation de la présentation
Le processus conduisant à la validation de la facilité d'utilisation	Le contrôle de la qualité	Les traductions
La gestion des demandes de modification des épreuves d'examen	L'évolution vers une évaluation sur ordinateur	

Comment l'IB révise-t-il la notation des enseignants ?

Public visé : les enseignants et les interlocuteurs désireux de comprendre pourquoi l'IB inclut les évaluations des enseignants dans ses processus, mais n'accepte pas toujours les jugements émis par ces enseignants.

Exemples de questions courantes :

- Pourquoi mes notes sont-elles modifiées ?
- Pourquoi le facteur de révision de notation change-t-il chaque année ?

Vous pourrez consulter une liste de sections susceptibles de présenter un intérêt particulier pour vous en cliquant ici.

En quoi consiste l'évaluation ?	Qu'est-ce que « l'évaluation » ?	Évaluation formative, sommative ou en tant que forme d'apprentissage... Pourquoi évaluer ?
Effet de remous et apprentissage	Notes et notes finales sont deux choses différentes	Révision de notation
Qu'est-ce que la révision de notation ?	La sélection des travaux de candidats	Les travaux inhabituels et « atypiques »

L'impossibilité de trouver un facteur de révision de notation	Échantillonnage dynamique	La hiérarchie de la révision de notation selon l'ancien système
Les commentaires sur l'évaluation interne	Établissement de la norme : comment préparer les examinateurs à la notation	La réunion de normalisation
Les copies d'entraînement	Les copies de validation	Les copies de contrôle
La marge de tolérance		

Quel est le sens de la formule « Des évaluations de qualité à l'ère du numérique » ?

La technologie s'imisce de plus en plus dans les activités quotidiennes. Pour dialoguer avec quelqu'un, il est devenu plus fréquent de passer par des formes de communication numérique, comme la messagerie disponible sur les médias sociaux, que d'écrire des lettres, et dans de nombreux secteurs professionnels, l'ordinateur constitue un outil incontournable pour exécuter le travail. L'informatique et la technologie au sens large ont également des effets positifs sur l'enseignement et l'apprentissage, et l'IB s'engage à aider les établissements scolaires qui proposent ses programmes à en faire le meilleur usage possible en classe.

Pour de nombreux élèves, la rédaction des dissertations et les recherches se font maintenant sur ordinateur, et écrire sur du papier pendant deux ou trois heures est devenu quelque chose d'inhabituel. Les examens de l'IB doivent être l'occasion pour les candidats de montrer ce qu'ils comprennent et non constituer une expérience unique qu'ils doivent apprendre à maîtriser.

En incluant la formule « Des évaluations de qualité à l'ère du numérique », le titre de la présente ressource reconnaît que pour garder leur pertinence, les évaluations de l'IB doivent se dérouler dans des conditions que les élèves connaissent bien et dans lesquelles ils se sentent à l'aise. Qui plus est, à l'heure de la transition, il est considéré comme impératif de rester fidèle aux principes fondamentaux de l'IB. Cela signifie que la technologie s'adapte aux besoins de l'évaluation, et jamais l'inverse.

Appréhender l'évaluation à l'ère du numérique permet surtout de veiller à rendre les évaluations toujours plus efficaces. En effet, il ne s'agit pas simplement d'éviter les inconvénients que pourrait présenter le passage du mode classique, à savoir les examens et les évaluations se déroulant en classe (évaluations internes), au format électronique, mais de se servir de la technologie pour dépasser les limites auxquelles l'IB s'est heurté au cours des 50 dernières années parce que tous les examens et travaux étaient effectués sur papier.

Mettre la technologie au service de l'évaluation signifie aussi avoir la possibilité d'augmenter la **validité des évaluations**, et la formule incluse dans le titre souligne l'enthousiasme de l'IB pour ce processus, tout en renforçant son engagement envers les principes qui sous-tendent toutes ses évaluations.

Quel est le lien entre cette ressource et les autres publications de l'IB ?

Le présent document vient rejoindre les autres ressources publiées par l'IB pour expliquer son approche de l'éducation. Il peut être compliqué de comprendre comment ces différentes ressources fondamentales sont liées les unes aux autres.

Vous trouverez ci-après un schéma illustrant ce lien et un tableau exposant les différents objectifs des principales ressources de l'IB.

Figure 2

Lien entre les principales publications de l'IB



Principales publications de l'IB couvrant tous les programmes

<i>Principes et pratiques de l'évaluation – Des évaluations de qualité à l'ère du numérique</i>	Cette publication explique l'approche globale adoptée par l'IB en matière d'évaluation et les mesures prises pour la mettre en pratique. Elle se concentre sur l'évaluation sommative (les évaluations officielles) dans le système éducatif de l'IB.
<i>En quoi consiste le système éducatif de l'IB ?</i>	Ce document vise à présenter clairement les éléments qui se trouvent au cœur du système éducatif de l'IB, en expliquant les idéaux sur lesquels tous les programmes de l'IB reposent. Grâce à sa description de la philosophie pédagogique de l'IB, il sert également de guide aux établissements scolaires tout au long de leur parcours, qui

Principales publications de l'IB couvrant tous les programmes	
	commence dès leur adhésion aux principes de l'IB et couvre les nombreuses années qu'ils connaîtront en tant qu'écoles du monde de l'IB.
<i>Normes de mise en œuvre des programmes et applications concrètes</i>	Cette publication fournit un ensemble de critères à l'aide desquels les établissements scolaires et l'IB peuvent évaluer le succès de la mise en œuvre des quatre programmes. Elle contient les normes de mise en œuvre des programmes (les conditions générales à remplir par les établissements pour mettre en œuvre tout programme de l'IB), leurs applications concrètes (les descriptions plus détaillées des normes de mise en œuvre communes à tous les programmes) ainsi que les exigences (qui sont spécifiques à chaque programme).

Publications de l'IB spécifiques à chaque programme	
<i>Des principes à la pratique</i>	Cette publication traite de l'enseignement et de l'apprentissage dans le cadre d'un programme de l'IB en particulier et explique les exigences de ce programme.
Guides pédagogiques (Programme du diplôme et PEI uniquement)	Chaque guide contient des informations détaillées sur une matière, un cours ou une discipline, comme la géographie ou les arts visuels. Les objectifs globaux, les objectifs spécifiques, le programme et les critères d'évaluation interne définis pour la matière en question y sont précisés. En règle générale, ces guides fournissent également des conseils supplémentaires pour l'enseignement et l'apprentissage spécifiques à la matière concernée.
Documents sur le contenu et l'enchaînement des programmes (PP uniquement)	En s'appuyant sur des exemples, cette série de documents explique comment consigner les attentes pédagogiques correspondant à chaque discipline. L'objectif est d'apporter un outil servant à informer l'ensemble de la communauté scolaire sur l'enseignement et l'apprentissage dans chaque discipline, de rendre transparents les éléments essentiels du PP en les inscrivant dans le contexte des différentes disciplines, et de préciser le rôle des disciplines dans un programme transdisciplinaire.
<i>Procédures d'évaluation</i>	Ce document expose l'ensemble des règles, règlements et processus spécifiques qui doivent être suivis pour procéder à l'évaluation propre à chaque programme de l'IB.
Règlement général	Il s'agit d'un regroupement des règles qui doivent être suivies par les écoles du monde de l'IB.
<i>Déroulement des examens du Programme du diplôme de l'IB (le POP utilise la version du Programme du</i>	Ces documents informent les coordonnateurs et les surveillants des règlements relatifs à l'administration et à la conduite des examens du programme. Un

Quel est le lien entre cette ressource et les autres publications de l'IB ?

Publications de l'IB spécifiques à chaque programme	
diplôme) et <i>Déroulement des examens sur ordinateur du Programme d'éducation intermédiaire de l'IB</i>	exemplaire de ces documents doit se trouver dans chaque salle d'examen.

Outre ces ressources précieuses, l'IB propose de nombreux autres documents qui fournissent des conseils ciblés aux enseignants, aux coordonnateurs et aux autres interlocuteurs, à partir des principes énoncés dans les principales publications susmentionnées. Ces documents sont disponibles sur le site Web de l'IB et sur le Centre de ressources pédagogiques.

Une évaluation s'appuyant sur la technologie

- Les élèves font une utilisation très large de la technologie. Ils s'en servent par exemple dans leur vie sociale, mais aussi pour leur apprentissage et la rédaction de leurs dissertations. Pourquoi leur demande-t-on encore de rédiger leurs examens à la main ?
- L'IB utilisera la technologie pour améliorer la qualité de ses évaluations et les rendre plus pertinentes. La technologie n'influencera pas les approches de l'évaluation qui seront choisies.
- Il est important de faire une distinction entre les effets de la technologie permettant d'apporter un soutien à des examinateurs spécialisés (notation électronique) et l'utilisation de la technologie permettant de concevoir des évaluations pertinentes pour les élèves (évaluation sur ordinateur – à la fois pour les examens et pour l'évaluation interne – et portfolios électroniques).
- Il ne faut pas confondre les expressions « sur ordinateur » et « en ligne » : aucune connexion à Internet ne sera requise pour pouvoir passer les évaluations de l'IB. Cette connexion ne sera nécessaire qu'une fois les évaluations terminées, pour envoyer les fichiers électroniques correspondants à l'IB.
- Les élèves devront avoir appris en classe à utiliser la technologie qui servira pour l'évaluation. En effet, il sera généralement plus facile de les habituer à cette technologie au moyen d'une méthode pédagogique que directement lors d'une évaluation.

À quoi ressemble l'évaluation sur ordinateur ?

Comme son nom l'indique, l'évaluation sur ordinateur désigne tout simplement une évaluation effectuée sur un ordinateur. Cela n'implique pas de choix multiple ni de fonction glisser-déposer (à moins que la situation ne s'y prête). Dans un premier temps, l'idée est de se servir des ordinateurs tout en conservant le style des épreuves d'examen actuelles. Par la suite, des améliorations seront apportées aux évaluations en incluant des questions interactives qui donneront plus d'authenticité aux examens ou les rendront plus accessibles aux élèves. Le potentiel offert par les évaluations sur ordinateur a déjà commencé à être exploré dans le PEI, mais une adoption plus lente a été choisie pour le Programme du diplôme. La vidéo ci-dessous présente des exemples de questions susceptibles d'être posées dans les évaluations sur ordinateur.

L'évaluation électronique du Programme d'évaluation intermédiaire (PEI) de l'IB

D'un point de vue pratique, passer des évaluations sur ordinateur signifie que chaque candidat doit disposer d'un ordinateur pour l'examen. Aucune connexion à Internet n'est cependant requise, étant donné que les épreuves sont chargées sur les ordinateurs avant le début des examens. Avec le modèle actuellement utilisé pour le PEI, l'ordinateur est « verrouillé » pendant toute la durée d'utilisation du logiciel d'examen, et des sauvegardes sont effectuées à la moindre interruption (le surveillant peut procéder à un redémarrage en cas d'interruption justifiée de l'examen).

L'évaluation sur ordinateur présente des avantages considérables : elle permet de tester ce que l'on souhaite véritablement tester et pas simplement ce qui est possible avec un examen sur papier. Tout le détail figure dans la section portant sur les avantages pour la validité, mais pour n'en citer que quelques exemples simples, ce système permet d'utiliser des supports vidéo, offre aux candidats la possibilité de faire pivoter des diagrammes et de les déplacer, mais aussi d'utiliser les outils de traitement de texte courants pour rédiger leurs dissertations. Dans la section « [L'équité pour tous : répondre aux besoins des candidats](#) », il est expliqué que l'évaluation sur ordinateur peut renforcer l'accessibilité, en donnant aux candidats la possibilité de modifier les couleurs et la taille de la police ainsi que de recourir à des outils comme des lecteurs d'écran.

Enfin, l'évaluation sur ordinateur ne sera pas statique. Les évaluations pourront en effet évoluer avec les avancées technologiques. À l'heure actuelle, seule l'utilisation des claviers et des souris ou des pavés

tactiles est prise en compte, mais dans quelques années, il est fort probable que les écrans tactiles seront monnaie courante et ouvriront d'autres horizons. En poussant la réflexion encore plus loin, quelle pourrait être la place de la réalité virtuelle dans certaines évaluations ? Un tel environnement pourrait donner encore plus d'authenticité aux évaluations en langues étrangères, et les films de science-fiction sont un cadre favorable au développement d'idées sur la façon de gérer des tâches techniques (questions scientifiques) dans des laboratoires de réalité virtuelle.

Un enseignement s'appuyant sur la technologie

La technologie est en train de modifier notre façon d'enseigner : certaines classes utilisent des tableaux interactifs, et des cours en ligne ouverts à tous mettent l'enseignement à la portée d'un plus grand nombre. La plupart des élèves ne rédigent plus leurs dissertations à la main, mais à l'aide d'un ordinateur.

L'abandon de l'écriture manuscrite au profit des ordinateurs crée en réalité un décalage inquiétant entre les pratiques de l'évaluation et les pratiques d'enseignement. Pour beaucoup d'élèves, l'expérience consistant à écrire au stylo deux ou trois heures de suite constitue une expérience presque exclusivement réservée au passage des examens, ce qui altère considérablement l'authenticité des évaluations.

Par ailleurs, en introduisant les évaluations sur ordinateur, il faut s'assurer que les candidats connaissent bien le style des évaluations, de façon à ce que l'utilisation de l'interface ne soit pas un obstacle risquant de masquer leur compréhension de la matière. Si cette difficulté peut être surmontée en fournissant aux élèves des outils de familiarisation et en leur faisant passer des examens d'entraînement, la meilleure solution reste que les élèves utilisent ces outils dans le cadre de l'enseignement.

Qu'est-ce que la notation électronique ?

L'expression « notation électronique » est employée pour désigner le système de notation des travaux des candidats se servant d'ordinateurs pour afficher les copies et consigner les notes. Sauf exception, les examinateurs ne reçoivent plus les copies d'examen ni les travaux d'évaluation interne au format papier : l'ensemble est numérisé pour que les examinateurs puissent accéder au format numérique en se connectant à Internet et en utilisant un logiciel de notation spécialisé.

Figure 3

Des examinateurs de l'IB utilisant le logiciel de notation électronique pour passer en revue le travail des candidats pendant une réunion de délibérations



Ce système présente des avantages de taille en ce qui concerne la vitesse et la qualité de la notation.

- L'IB garde le travail des candidats en sa possession. De ce fait, si ce travail doit être revu, pour quelque raison que ce soit, par un second examinateur, il est alors immédiatement disponible et non soumis aux délais posés par une expédition dans un autre pays.
- Le travail des candidats peut être rendu anonyme afin de réduire les risques de biais dans la notation de l'examineur.
- Un processus de contrôle de la qualité plus rigoureux peut être mis en œuvre avec la vérification des normes de notation au fil de la correction.
- Les travaux des candidats d'un même établissement peuvent être répartis entre tous les examinateurs, au lieu d'être tous notés par un seul examinateur. Le contrôle de la qualité peut ainsi être poussé encore plus loin.

L'IB a commencé à utiliser la notation électronique en 2010, et ce système n'impose pas aux candidats de rendre un travail réalisé à l'aide d'un ordinateur. À l'heure actuelle, les réponses aux épreuves d'examen sont, pour la plupart, rédigées à la main, puis les copies sont numérisées pour être utilisables au format numérique.

Les risques à éviter avec l'évaluation sur ordinateur

Développer une évaluation sur ordinateur demande tout d'abord de prendre conscience d'un certain nombre de difficultés importantes, puis de les surmonter pour assurer une équité entre tous les candidats et tous les établissements. Quelques-uns des principaux risques à gérer sont énumérés dans la liste qui suit. Une analyse plus poussée sera réalisée dans plusieurs sections du présent document ainsi que dans d'autres publications de l'IB traitant de l'évaluation électronique.

La charge pesant sur les établissements

Force est de reconnaître que l'utilisation de la technologie est à présent parfaitement intégrée dans les gestes quotidiens des élèves. Par conséquent, leur demander d'écrire sur du papier serait une exigence anachronique. Il n'en demeure pas moins que les établissements ont besoin de temps pour modifier leurs pratiques et que les évaluations sur ordinateur nécessitent des aménagements particuliers par rapport aux examens sur papier. L'introduction des évaluations sur ordinateur se fera avec mesure au niveau du Programme du diplôme et du POP, afin que tous les établissements puissent suivre les avancées technologiques, mais aussi pour qu'ils n'aient pas à s'adapter de manière trop brutale, quelle que soit leur situation géographique.

Le risque de défaillance technologique

Aucun degré de défaillance technologique ne peut être accepté quand il s'agit du passage des examens. En effet, chaque candidat est en droit de vivre cet événement sans heurts. Pour y parvenir, deux principes sont suivis. Premièrement, l'évaluation peut être passée sur un seul ordinateur et aucune connexion à Internet n'est requise. Deuxièmement, les établissements ont la possibilité de suivre un processus de test complet ou de conformité la veille des examens. L'expérience montre en effet que la plupart des problèmes viennent de l'infrastructure ou de certains paramètres des établissements, ce qui peut donc être repéré et résolu bien avant les examens.

Les questions de sécurité

Le niveau de sécurité des évaluations sur ordinateur doit être au moins identique à celui des examens sur papier. Ces évaluations sont conçues de façon à « verrouiller » toutes les autres fonctionnalités de l'ordinateur pendant leur déroulement. Elles sont par ailleurs cryptées et protégées par un mot de passe.

Même si la sécurité demande à être constamment renforcée, en raison des développements de plus en plus sophistiqués qui voient le jour, il ne faut pas perdre de vue que les examens sur papier utilisés actuellement sont eux aussi exposés à certains risques.

La technologie dans l'intérêt de la technologie

L'IB s'est engagé à utiliser la technologie pour améliorer la qualité des éléments pouvant être évalués, et non pour le simple fait de posséder cette technologie. Ainsi, l'IB continuera à expliquer clairement comment il apporte une valeur ajoutée à la validité de ses évaluations, et le processus d'élaboration des épreuves reflétera ce progrès pour éviter toute utilisation de la technologie à tout prix.

Le préjudice subi par certains groupes de candidats

Force est de reconnaître que certains candidats réussiront mieux que d'autres parce qu'ils pourront passer les évaluations en se servant d'ordinateurs. Il ne faut cependant pas perdre de vue que les examens sur papier peuvent également entraîner un préjudice : les candidats ayant des difficultés à écrire pendant longtemps au stylo ou écrivant lentement sont déjà désavantagés.

Par principe, dans aucun des deux cas (sur papier ou sur ordinateur) l'objectif n'est de mesurer la capacité des candidats à utiliser un stylo ou un ordinateur, et les tâches d'évaluation ne seront pas conçues de façon à procurer un avantage aux personnes capables de taper ou d'écrire rapidement.

Afin de garantir la bonne application de ce principe, l'IB accorde une importance particulière aux études publiées sur les « effets de la technologie ».

L'évolution des normes, vers le haut ou vers le bas

Les notes finales de l'IB ont un sens, et ce sens sera protégé par le jugement spécialisé des examinateurs superviseurs, qui s'appuie sur les données sur les résultats. À titre d'exemple, il peut être plus facile d'élaborer une dissertation mieux structurée en ayant la possibilité de couper et de coller des paragraphes au fil de la rédaction. Une attention particulière sera portée à cet élément lors de l'établissement des seuils d'attribution des notes finales. De cette façon, un candidat autant à l'aise avec le travail sur papier qu'avec le travail sur ordinateur obtiendra le même résultat, qui rendra compte de sa compréhension et de sa capacité d'analyse dans la matière.

L'évaluation sur ordinateur permet de tester de nouveaux aspects ainsi que des aspects qui avaient toujours été valorisés, mais qu'il était impossible d'évaluer jusque-là. De ce fait, il est possible que les candidats aient à montrer d'autres éléments dans les évaluations (voir la section « Définition des normes »). Ce changement sera toutefois opéré dans le respect des objectifs de longue date de l'IB et de façon à refléter les qualités décrites dans les descripteurs des notes finales en vigueur.

Les obstacles empêchant les établissements de proposer les programmes de l'IB

En ce qui concerne la charge pesant sur les établissements, un travail sera réalisé avec la communauté de l'IB pour que le calendrier d'introduction des évaluations sur ordinateur permette à tous les établissements intéressés par une technologie offrant une évaluation de meilleure qualité d'en bénéficier. L'IB estime que le potentiel de la technologie doit être exploité intelligemment pour améliorer la qualité de l'enseignement, de l'apprentissage et, par conséquent, de l'évaluation.

En quoi consistent les principes et les pratiques de l'évaluation ?

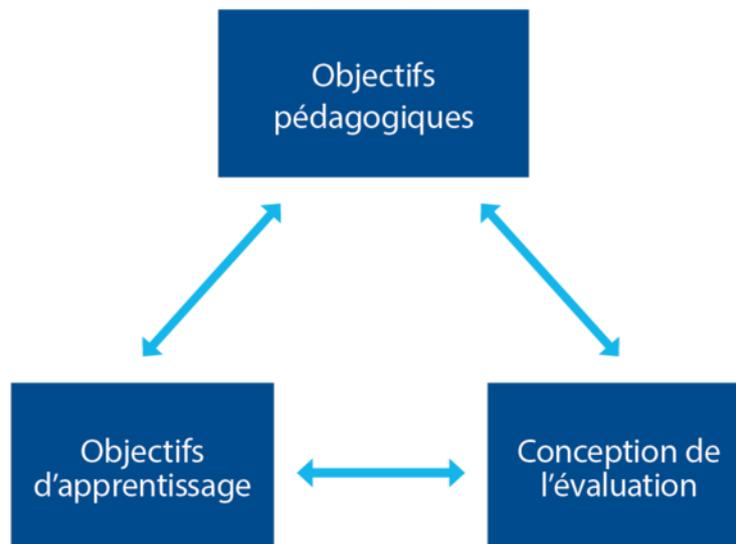
Les **principes de l'évaluation** sont les propositions de base jugées importantes pour la conception, la mise en œuvre, la correction et la notation des évaluations menant aux diplômes et certificats. Ils découlent des éléments essentiels du système éducatif de l'IB, le principe prépondérant étant que les évaluations doivent être au service de l'éducation, et non venir la fausser.

Les **pratiques de l'évaluation** sont la transposition efficace et concrète des principes. Elles tiennent compte des exigences contradictoires et des limites pratiques posées par une utilisation à travers le monde, tout en respectant l'intégrité, une idée phare de la philosophie de l'IB.

Les pratiques de l'évaluation sont décrites de manière très générale, et elles doivent être mises en œuvre de façon adéquate, en fonction du contexte propre à chaque matière, discipline et programme. Elles expliquent de façon générale ce qui doit être fait, sans donner de détails précis sur la marche à suivre au quotidien ni sur les applications différentes en fonction des matières concernées.

Figure 4

Lien entre les objectifs pédagogiques, les objectifs d'apprentissage et la conception de l'évaluation



Terminologie de l'évaluation

- Comme la plupart des domaines, l'évaluation possède une terminologie spécifique et un ensemble d'expressions qu'il est utile d'apprendre à connaître pour pouvoir s'exprimer sur ce sujet.
- L'introduction de l'évaluation sur ordinateur conduira également à une modification de certains termes utilisés.

Quels sont les principaux termes à comprendre ?

Comme de nombreux autres domaines, y compris l'enseignement, l'évaluation a une phraséologie spécifique. Il est évident que cet aspect peut poser des problèmes pour entamer une discussion sur l'évaluation ; c'est pourquoi l'IB s'est engagé à s'exprimer dans un langage simple et à faire un usage aussi limité que possible du jargon.

En dépit de cet effort de simplification, certains termes et concepts spécifiques restent incontournables, et il est utile d'en saisir le sens avant de lire le présent document.

Candidat et élève

Le terme « candidat » désigne toute personne passant une évaluation de l'IB. Le terme « élève » désigne toute personne étudiant des cours de l'IB. Les élèves deviennent donc des candidats lorsqu'ils passent une évaluation de l'IB.

En règle générale, il est important de faire une distinction entre ces deux notions uniquement pour le traitement des données, lorsque les chiffres ou les résultats peuvent être différents entre les deux groupes.

Évaluation, examen, test et composante

Le terme « évaluation » désigne toute tâche effectuée par des candidats pour montrer leurs capacités. Le terme « examen » désigne une forme d'évaluation particulière, au cours de laquelle les candidats doivent répondre à des questions formulées par l'IB, et ce, dans des conditions strictement contrôlées.

Dans le présent document, le terme « test » est utilisé au sens générique pour désigner un examen, même si les articles de recherche sur l'évaluation peuvent lui conférer un sens spécifique.

Parfois, l'évaluation globale d'un candidat est divisée en plusieurs éléments distincts, passés à des moments différents. Chaque élément porte alors le nom de « composante d'évaluation » ou de « composante » tout simplement. Parmi les exemples de composantes, il est possible de citer les épreuves orales, l'épreuve (d'examen) 1, l'épreuve (d'examen) 2 ou l'évaluation interne.

Notation et attribution d'une note finale

Des notes sont attribuées pour indiquer dans quelle mesure le candidat a correctement répondu à une question. Le nombre de points pouvant être attribué varie en fonction des questions et des examens. Une note finale décrit la qualité de l'accomplissement du candidat, et elle doit signifier la même chose pour chaque examen, année et matière. Il s'agit d'une différence de taille.

De plus amples informations sont fournies à ce sujet dans la section « [Quelle est la différence entre les notes et les notes finales ?](#) ».

Question et élément

Le terme « question » est l'expression générale utilisée pour désigner une tâche bien distincte, qui peut être composée de plusieurs parties amenant le candidat à traiter un problème.

Le terme « élément », ou « élément de question », désigne chaque réponse apportée par un candidat. Ainsi, la partie c de la question 1 [1c] ou la section iii de la partie e de la question 7 [7(e)(iii)] seraient des éléments.

Auteur d'épreuve et examinateur

Le terme « auteur d'épreuve » désigne toute personne élaborant les questions et le barème de notation associé qui seront utilisés pour les évaluations. Les auteurs d'épreuves sont presque toujours des examinateurs superviseurs, bien que ce travail requière des compétences différentes et qu'il ne soit pas forcément lié au rôle d'examineur.

Le terme « examinateur » désigne toute personne notant le travail d'un candidat. Les équipes d'examineurs sont organisées selon une hiérarchie précise, les deux fonctions les plus élevées étant celles d'examineur principal (pour chaque composante) et d'examineur en chef (pour une matière).

Les examinateurs superviseurs, dirigés par leur examinateur principal et leur examinateur en chef, ont pour responsabilité de recommander où établir les seuils d'attribution des notes finales. Cet établissement des seuils doit être décidé une fois la notation terminée.

Le détail des responsabilités associées aux différentes fonctions est donné dans la section « [Processus d'évaluation – Rôles et responsabilités](#) ».

Les nouveaux termes pour l'évaluation électronique

Le passage à l'évaluation électronique signifie que l'IB et la communauté de l'évaluation au sens large doivent commencer à imaginer de nouveaux termes pour décrire leurs pratiques. Même s'il est parfois possible de suivre la tendance et d'ajouter l'adjectif « électronique » aux termes existants (par exemple, utiliser « épreuve électronique » pour désigner la version électronique d'une épreuve d'examen), d'autres termes sont très différents de leur équivalent dans la version papier.

Dans la liste qui suit, seule l'expression « évaluation sur ordinateur » est largement utilisée dans le présent document, mais vous pourrez trouver les autres termes si vous effectuez d'autres lectures sur l'évaluation électronique de l'IB.

Évaluation sur ordinateur

L'expression « évaluation sur ordinateur » désigne toute évaluation se déroulant sur un ordinateur. Elle englobe l'idée des examens passés sur un ordinateur, mais aussi le travail d'évaluation réalisé en classe s'il est essentiellement réalisé dans un environnement informatique.

La différence fondamentale entre une évaluation sur ordinateur et l'utilisation d'un ordinateur en classe est que le premier cas s'appuie sur un logiciel strictement contrôlé pour rassembler les preuves de l'évaluation. Par exemple, les candidats peuvent être autorisés à aller sur Internet dans le cadre de leur étude d'un thème, mais avec l'évaluation sur ordinateur, il est probable que cela servira à guider la recherche ou à consigner ce que les candidats ont consulté, de façon à pouvoir évaluer leur manière de faire des recherches sur le Web.

Le terme « sur ordinateur » a été choisi avec soin. Il ne s'agit pas d'un dispositif en ligne et, plus précisément, aucune connexion à Internet n'est requise pour que les candidats puissent passer les évaluations sur ordinateur de l'IB.

Fichier de réponses et réponse du candidat

Dans l'univers du papier, l'expression « copie du candidat » est utilisée. Cependant, avec les multiples possibilités offertes par l'évaluation électronique, le travail des candidats pourra également être multimédia. Le processus menant à la production du travail fait partie de ce qui est évalué. Par conséquent, le résultat d'une évaluation électronique porte le nom de « réponse du candidat ».

L'expression « fichier de réponses » désigne tout simplement le fichier informatique qui contient la réponse du candidat. Ce sont ces fichiers qui doivent être envoyés à l'IB.

Outil de familiarisation

Les candidats doivent impérativement savoir utiliser efficacement un ordinateur pour ne pas être défavorisés lors des examens. L'outil de familiarisation est un exemple vierge d'évaluation sur ordinateur dans lequel les candidats peuvent s'entraîner à utiliser tous les paramètres existants.

Figure 5



Les candidats devront également passer des examens d'entraînement sur le modèle de l'évaluation sur ordinateur, sous contrainte de temps et en prenant au sérieux leurs réponses aux questions. Pour qu'ils puissent s'habituer au fonctionnement du logiciel, les candidats peuvent être autorisés à utiliser l'outil de familiarisation autant qu'ils le souhaitent.

En cas de doute sur le sens d'un terme

Il est évident que les évaluations de l'IB s'accompagnent d'un grand nombre de termes qui peuvent être méconnus des enseignants et des élèves. Dans chacun des documents de l'IB, l'idée est d'éviter toute ambiguïté tout en conservant des propos précis.

Dans ce but, l'IB utilisera les abréviations avec parcimonie, en veillant à ce qu'elles soient expliquées la première fois qu'elles apparaissent dans une section. De même, il sera fait en sorte que les termes importants soient expliqués la première fois qu'ils sont utilisés, soit directement dans le document soit par l'intermédiaire de liens renvoyant vers des explications.

Si vous rencontrez des termes inconnus dans le présent document, veuillez vous référer au [glossaire](#) qui contient une liste complète de termes liés à l'évaluation.

En quoi consiste l'évaluation ?

- L'évaluation englobe un grand nombre de choses très différentes.
- Dans le domaine de l'éducation, l'évaluation comporte généralement un volet formatif (évaluation au service de l'apprentissage) et un volet sommatif (évaluation de l'apprentissage). La tendance actuelle va dans le sens d'une « évaluation en tant que forme d'apprentissage ».
- L'évaluation doit être conçue avec soin pour pouvoir atteindre les objectifs pour lesquels ses résultats sont utilisés. Un excellent test formatif peut être très mauvais pour effectuer des mesures sommatives.
- L'évaluation peut influencer les pratiques d'enseignement et doit être conçue de façon à ce que tout « effet de remous » soit positif.

Qu'est-ce que « l'évaluation » ?

Le terme « évaluation » peut désigner toutes les façons de recueillir et d'évaluer des données sur l'accomplissement d'un candidat. Parmi les évaluations courantes, il est possible de citer les tests, les examens, les travaux pratiques approfondis, les projets, les dossiers et les travaux oraux. Dans certaines situations, les évaluations se déroulent sur une longue période et, dans d'autres, elles prennent quelques heures. Les évaluations seront parfois jugées par l'enseignant du candidat et, parfois, le travail sera évalué par un examinateur externe.

Vous pourrez remarquer que les verbes « juger » et « évaluer » ont été préférés aux verbes « noter » et « attribuer une note finale ». Cela s'explique par le fait qu'une distinction importante est faite entre ces deux concepts, comme cela sera indiqué par la suite.

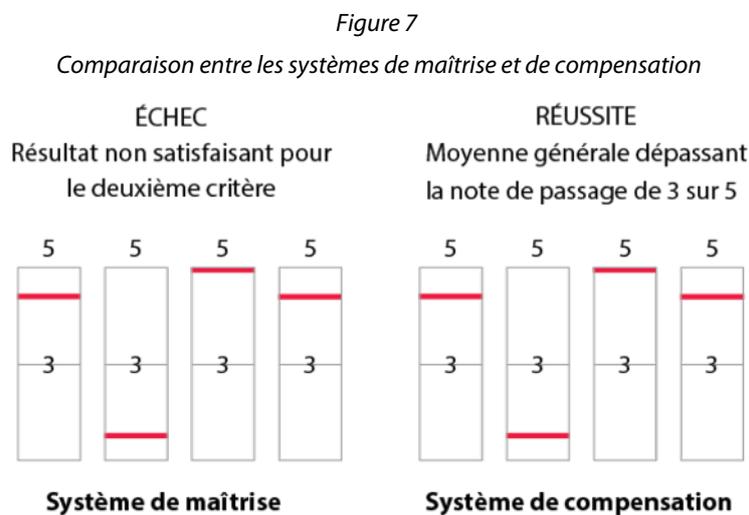
Figure 6

Exemples de différents types d'évaluations



Les tâches d'évaluation qui vérifient le niveau de compétence d'un candidat s'appuient soit sur un système de **compensation**, soit sur un système de **maîtrise**. Le système de compensation est utilisé dans la plupart des examens externes. Il permet à un excellent résultat dans un domaine d'atténuer un mauvais résultat dans un autre. Par exemple, imaginez qu'un examen soit composé de 3 questions, chacune notée sur 10. La note de passage est 15, et elle pourrait être obtenue par un candidat avec les notes 5, 5 et 5 (s'il réussissait à peu près la moitié de chaque question) ou par un candidat avec les notes 10, 3 et 2. Dans le second cas, un résultat parfait à la première question vient compenser les mauvais résultats obtenus aux deux autres questions.

À l'inverse, le système de maîtrise exige un niveau minimal (maîtrise) dans chaque partie d'une évaluation. En reprenant l'exemple précédent, si un 5 était exigé pour chaque question, le premier candidat aurait réussi l'évaluation alors que le second aurait échoué, tout comme un troisième candidat qui aurait obtenu 10, 10 et 4, soit un total de 24 points, mais pas les 5 points exigés pour la troisième question.



Les évaluations suivant le système de maîtrise sont généralement utilisées dans des contextes plus professionnels (milieu du travail) où il n'est pas adapté d'être très bon dans un domaine, mais mauvais dans un autre. Par exemple, pour confectionner une robe, aucun niveau élevé d'expertise en matière de stylisme ne peut compenser l'absence de compétences de base en couture.

L'IB s'appuie sur toute une gamme d'outils d'évaluation, notamment des épreuves d'examen destinées à être passées à la fin du programme, ainsi que sur une multitude de tâches d'évaluation (essais, mémoires de recherche, travaux écrits, entretiens oraux, recherches scientifiques et mathématiques, rapports sur le travail de terrain et représentations artistiques) portant sur des matières très diverses et effectuées par les candidats à différents moments et dans des conditions variables au cours du programme.

Évaluation formative, sommative ou en tant que forme d'apprentissage... Pourquoi évaluer ?

L'évaluation peut être utilisée à diverses fins. L'objectif visé par une évaluation donnée aura une grande influence sur la manière dont elle est conçue. Deux raisons sont habituellement avancées pour procéder à une évaluation : des fins formatives et sommatives.

Avec l'évaluation formative, le but est de fournir un retour d'information détaillé aux enseignants et aux élèves sur la nature des points forts et des points faibles des élèves, et d'aider au développement des capacités. Les évaluations de type interaction directe, comme une discussion, entre l'enseignant et l'élève sont alors particulièrement utiles.

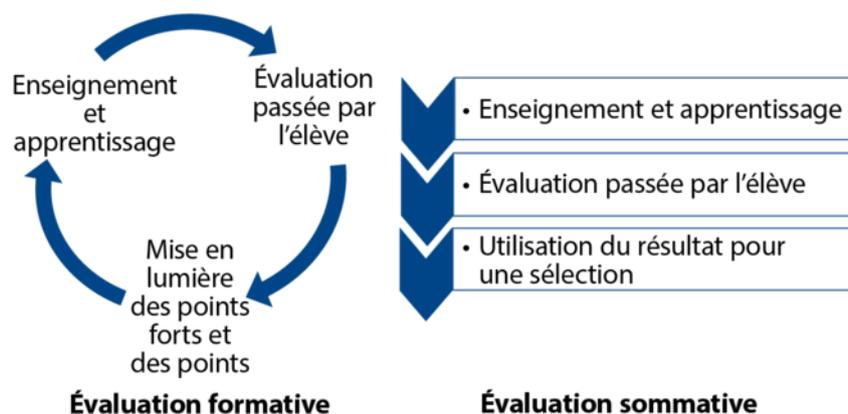
Vygotsky (1962) voit l'enseignant comme un soutien plutôt que comme quelqu'un qui dirige l'enseignement, et l'enseignant doit donc utiliser des tâches et des outils d'évaluation qui aident les élèves à travailler dans ce que Vygotsky appelle la « zone proximale de développement ». Il s'agit de l'écart (la gamme d'accomplissements) entre ce que l'élève peut faire seul et ce qu'il peut faire avec l'aide de l'enseignant.

Ce concept rappelle la notion d'« étayage » (ou soutien à l'apprentissage) de Wood *et al.* (1976). Durant le processus d'« étayage », l'enseignant fournit l'« étayage » autour duquel l'élève pourra « construire » son apprentissage, mais seul l'élève peut effectuer cette construction. L'objectif de l'enseignant doit être de préparer des évaluations formatives qui posent aux élèves des défis d'un niveau adéquat et de continuer à adapter ce niveau au fur et à mesure que les élèves progressent.

À l'inverse, l'évaluation sommative se concentre sur une mesure de **ce que** l'élève peut faire, afin de montrer la réussite d'un programme de formation ou le degré de préparation au passage à l'étape éducative suivante. Alors que l'évaluation formative s'intéresse aux **raisons** pour lesquelles un élève fait quelque chose, l'évaluation sommative cherche à savoir si l'élève a **correctement** fait les choses. Même si cela peut sembler moins utile que comprendre les raisons, il est intéressant d'étudier l'objectif particulier de l'évaluation sommative, qui consiste à émettre un jugement sur l'élève, et non à orienter l'enseignement à venir. Les personnes doutant de la nécessité d'une évaluation sommative sont invitées à lire la section « Description de la réussite : les résultats des candidats à l'évaluation sommative ».

Figure 8

Deux utilisations différentes de l'évaluation, selon la méthode formative ou sommative



Avec l'évaluation formative, il est plus important d'identifier correctement les connaissances, les compétences et la compréhension que les élèves n'ont pas encore développées, que de mesurer avec précision le niveau de chaque accomplissement de l'élève. L'équilibre entre ces concepts porte le nom de « validité », une notion expliquée plus en détail par la suite. Cet équilibre entre l'accomplissement de l'élève et la qualité du retour d'information est inversé avec l'évaluation sommative, dans laquelle les résultats de l'évaluation sont utilisés pour prendre des décisions concernant l'élève, souvent pour une sélection compétitive en vue d'un emploi ou de possibilités d'études, mais aussi pour mieux décider de la poursuite de l'enseignement.

Il est bon de noter que toute analyse des différents systèmes d'évaluation nationaux révélera une grande variété de techniques et de méthodes d'évaluation. Tous ces systèmes ont leurs forces et leurs faiblesses en ce qui concerne les techniques, les ressources et le temps ainsi que leur impact sur le système éducatif national auquel ils sont associés. Même s'il était possible, dans un contexte donné, de recommencer à zéro la conception d'un système d'évaluation, il n'existe pas de « meilleure technique » universelle qui pourrait être adoptée. Au contraire, les choix qui sont faits lors de la conception des systèmes d'évaluation reflètent inévitablement les valeurs et les priorités du contexte social plus vaste dans lequel ils sont faits. Pour approfondir ce sujet, veuillez vous référer à Cresswell (1996) et Broadfoot (1996).

Il convient également de noter que l'évaluation sommative est de plus en plus utilisée comme une mesure de la qualité de l'enseignement, et cela ajoute une dimension supplémentaire aux raisons pour lesquelles l'évaluation est réalisée, ce qui se fait davantage au profit du système éducatif qu'à celui de l'élève.

Effet de remous et apprentissage

Mesurer quelqu'un ne l'a jamais fait grandir plus vite.

(Meighan, 2004)

Nous avons besoin d'un processus d'évaluation qui soit aussi valide que possible. De plus, un tel processus ne doit pas, en raison de son effet de remous, fausser la qualité de l'enseignement, être trop lent ou absorber une trop grande quantité de nos ressources scolaires, qui sont peu abondantes.

(Peterson, 1971)

Dans cette citation, Alec Peterson, le fondateur de l'IB, reconnaît que la manière de mener l'évaluation à bien peut influencer la manière d'aborder l'apprentissage, et que cela présente un risque. Surgenor (2010) résume bien les raisons pour lesquelles cela peut avoir des effets positifs et négatifs.

Dans un article publié en 1971, Snyder propose que les élèves créent leur propre compréhension du programme d'études à partir de messages implicites et explicites sur les éléments qui comptent pour l'évaluation, ce qu'il décrit comme le « programme caché ». Cela peut les amener à comprendre comment être reçu à l'examen dans une matière, mais pas à comprendre la matière elle-même. Cette idée est bien résumée dans la citation qui suit.

Lorsque j'ai repassé l'examen, j'ai uniquement cherché à le réussir. J'ai obtenu un résultat de 96 %, et le type n'a pas compris pourquoi j'avais échoué la première fois. Je lui ai répondu que, cette fois, j'avais uniquement cherché à réussir l'examen et non à comprendre la matière. Étant donné que je ne comprends toujours pas cette matière, j'ai manqué mon objectif d'une certaine façon.

(Gibbs, 1992, p. 101)

Ce concept de remous est également résumé dans un adage souvent repris : « Si ce n'est pas évalué, ce ne sera pas enseigné. » Cela souligne l'idée que l'évaluation et l'enseignement ne peuvent pas être considérés comme deux choses indépendantes.

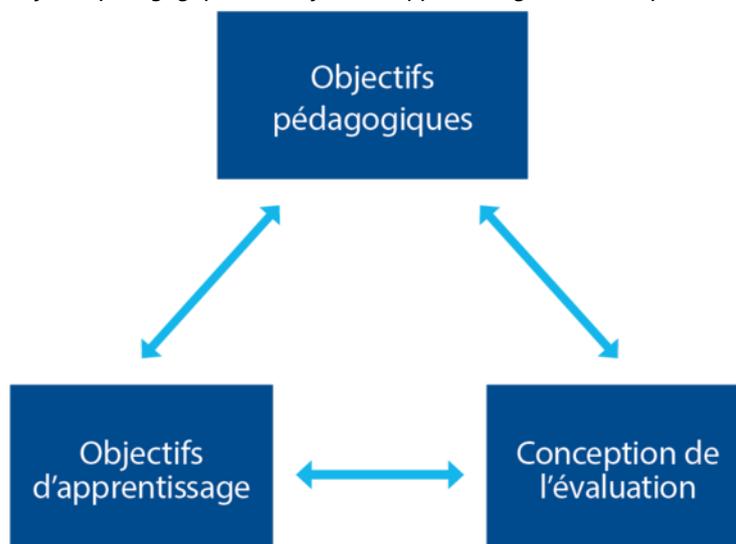
L'évaluation devrait servir à fournir des preuves de l'atteinte des objectifs pédagogiques par les candidats, mais cette fonction peut présenter plusieurs défaillances.

La plus vraisemblable est que les objectifs d'apprentissage sur lesquels repose l'évaluation ne soient pas un bon reflet du but premier de l'éducation. Cette idée pourrait être illustrée dans le milieu professionnel par un cours de formation pour sportifs de haut niveau ayant des objectifs d'apprentissage axés sur la compréhension des règles de l'athlétisme. Un autre problème récurrent est que, tandis que l'évaluation couvre tous les objectifs visés, l'essentiel du test se concentre sur des tâches faciles à évaluer mais peu importantes pour les objectifs pédagogiques globaux.

Cette interdépendance entre évaluation et but pédagogique peut être illustrée par le schéma ci-après (modèle de Furst [1958], présenté dans Frith et Macintosh, 1984). Tout élément qui ne serait pas relié à un autre impliquerait presque inévitablement une évaluation de mauvaise qualité.

Figure 9

Lien entre les objectifs pédagogiques, les objectifs d'apprentissage et la conception de l'évaluation



Comme il est expliqué dans la publication intitulée *En quoi consiste le système éducatif de l'IB ?*, une approche constructiviste de l'apprentissage a été adoptée. Celle-ci confère un rôle actif à l'élève et reconnaît l'importance du contexte dans l'efficacité de l'apprentissage (Murphy, 1999). Si l'évaluation vise à renforcer l'efficacité de l'enseignement et de l'apprentissage, alors elle doit être conçue selon la théorie de l'apprentissage constructiviste. Pour consulter des travaux de recherche sur ce concept, voir Black (1999), Shepard (1992), Wood (1998) ainsi que Lambert et Lines (2000).

L'évaluation formative montre le lien le plus immédiat avec la manière dont les élèves apprennent, ce qui lui vaut parfois la qualification d'évaluation *au service de* l'apprentissage. L'évaluation sommative est quant à elle parfois considérée comme une évaluation *de* l'apprentissage. Une telle approche minimise l'impact majeur qu'a l'évaluation sommative sur ce qui est réellement appris en classe. Toute évaluation doit apporter son soutien à un apprentissage approprié. L'évaluation sommative ne consiste pas uniquement en une activité succédant à l'apprentissage. Elle doit au contraire être conçue de manière à faire partie intégrante de l'enseignement et de l'apprentissage.

Notes et notes finales sont deux choses différentes

Pour bien conduire des évaluations sommatives et en tirer profit, il est important de comprendre la différence entre la notation du travail des candidats et l'attribution d'une note finale.

- Dans la notation, le travail produit par un candidat est récompensé au regard d'un barème de notation ou d'un autre cadre similaire. Il s'agit d'une indication de la proportion de la tâche d'évaluation que le candidat a réussie. La note en tant que telle n'a pas d'autre portée.
- En attribuant une note finale, l'examineur émet un jugement sur la qualité du travail du candidat au regard d'une norme définie et en prenant en compte la difficulté de la tâche ainsi que la proportion de la tâche qui a été réalisée. Par conséquent, la note finale a un sens ou une pertinence et doit en principe pouvoir être comparée aux résultats obtenus dans les autres évaluations.

Figure 10

La notation peut être comparée à la qualité (saveur) de la nourriture préparée, mais la note finale rend quant à elle compte de la complexité de la préparation culinaire.



Un candidat pourrait obtenir une note finale élevée en réussissant uniquement une proportion limitée d'une question très difficile, mais ne pas réussir à obtenir cette même note finale en répondant correctement à de nombreuses questions bien moins complexes.

Comme il est expliqué par la suite, il n'est pas nécessaire de faire des références aux aspects réussis par le candidat pour expliquer le niveau correspondant à la note finale, même si l'IB suit une telle approche. En effet, il existe d'autres systèmes tout à fait cohérents et très réputés, dans lesquels le niveau d'un candidat est déterminé par rapport au résultat de ses pairs.

Dans ses évaluations, l'IB se sert habituellement des notes pour avoir une idée des résultats globaux (système de compensation). Il s'intéresse ensuite à ce qui est exprimé par les candidats ayant obtenu ce nombre de points pour fixer une limite (seuil d'attribution des notes finales) au-delà de laquelle attribuer une certaine note finale aux candidats ayant obtenu plus de points. Ce processus est expliqué plus en détail dans la section « [Les pratiques de l'évaluation de l'IB](#) ».

Quelle est l'adéquation avec l'objectif visé ? Comprendre la validité

- S'intéresser à la validité signifie se demander si une évaluation est adaptée à l'objectif visé. Il s'agit d'une notion complexe, qui possède de multiples facettes.
- Les évaluations sont utilisées à diverses fins.
- Une évaluation peut être valide pour une fin, mais pas pour une autre. Par exemple, elle peut tester la capacité d'un candidat à orthographier les mots, mais pas son aisance d'expression.
- Il peut être considéré que ce sont les objectifs associés aux résultats de l'évaluation qui peuvent être valides ou non, plus que les évaluations elles-mêmes.
- La principale préoccupation de l'IB est de savoir si le programme est valide, puis si les éléments de ce programme (par exemple, les différents cours) sont valides et, enfin, si chaque évaluation est valide.

Qu'est-ce que la validité ?

Tout le monde est un génie. Mais si vous jugez un poisson sur sa capacité à grimper à un arbre, il va passer toute sa vie à penser qu'il est stupide.

(Anonyme, même si cette citation est souvent attribuée à Albert Einstein)

Pourquoi passer des examens ? À quoi servent les examens ? Ces questions simples en apparence se trouvent au cœur du sens donné à une évaluation « adaptée à l'objectif visé » ou « mesurant ce qu'il faut », et, bien souvent, les réponses apportées sont en contradiction les unes avec les autres. Par ailleurs, il peut arriver que tous les objectifs pédagogiques ne soient pas évalués par un examen ou que certains soient impossibles à évaluer.

À titre d'exemple, imaginez que les quatre possibilités suivantes soient données comme objectifs d'une évaluation en mathématiques :

- détermination de la compréhension développée par l'élève après le programme ;
- système de sélection en vue de la poursuite des études ou de l'entrée dans le monde du travail ;
- indication de la réussite future ;
- consolidation de l'enseignement des objectifs pédagogiques du programme.

Malgré leur simplicité, ces objectifs sont difficiles à concilier : si la réussite future en mathématiques dépend plus de l'analyse infinitésimale que de la géométrie, l'accent doit-il être mis sur l'analyse infinitésimale ? Quel est le lien entre cet objectif et le premier, qui consiste à déterminer ce que l'élève sait (en géométrie) ?

Ces quatre possibilités ne sont en rien définitives. Pour Newton (2007), les résultats de l'évaluation peuvent être utilisés à de nombreuses fins. Il en donne les exemples qui suivent.

- | | |
|---|---|
| <ul style="list-style-type: none">• Utilisation à des fins d'évaluation sociale• Utilisation à des fins formatives• Utilisation à des fins de contrôle de la progression de l'élève• Utilisation à des fins de transfert• Utilisation à des fins de répartition• Utilisation à des fins de diagnostic• Utilisation à des fins d'orientation• Utilisation à des fins de mesure de la compétence | <ul style="list-style-type: none">• Utilisation à des fins de permis• Utilisation à des fins de décisions scolaires• Utilisation à des fins de contrôle de l'établissement• Utilisation à des fins d'affectation des ressources• Utilisation à des fins d'intervention organisationnelle• Utilisation à des fins d'évaluation du programme |
|---|---|

- | | |
|---|---|
| <ul style="list-style-type: none">• Utilisation à des fins de sélection | <ul style="list-style-type: none">• Utilisation à des fins de contrôle du système• Utilisation à des fins de comparabilité |
|---|---|

Face à un tel éventail d'utilisations contradictoires, il pourrait être tentant de suggérer que l'évaluation et les notes finales des examens ne peuvent pas être adaptées à l'objectif visé et donc qu'il serait préférable d'y renoncer. Le problème avec une telle approche est que des personnes devront malgré tout prendre des décisions et que si les résultats de l'évaluation ne sont pas disponibles, alors d'autres paramètres – moins bien pensés et compris – risquent d'être utilisés pour cette prise de décisions. Il s'agit de l'argument défendu par Cresswell ci-après.

À l'évidence, si les autres critères sont moins fiables que les examens, compter davantage sur ces critères signifiera faire reposer la sélection sur des décisions moins fiables.

(Cresswell, 1986, p. 37 – 54)

Cette idée de diplôme ou certificat « adapté à l'objectif visé » est englobée dans la notion de « validité ».

Le terme « validité » revient souvent, mais les universitaires cherchent depuis longtemps à en donner une définition précise (Newton, 2012). Et pour compliquer les choses, il existe une notion plus étroite de la validité, souvent associée au terme « fiabilité », pour désigner la capacité d'une évaluation à mesurer réellement ce qu'elle est censée mesurer. Il s'agirait par exemple de ne pas recourir à un examen écrit pour montrer qu'un candidat sait faire un gâteau. Pour éviter toute confusion dans le présent document, l'expression « pertinence de construit » a été choisie pour évoquer ce second concept. Pour obtenir de plus amples informations, veuillez consulter la section « [Pertinence de construit et authenticité](#) ».

En fin de compte, la validité est une question de jugement. Les exigences contradictoires de l'évaluation qui ont été rapidement évoquées ci-dessus seront traitées plus en détail par la suite, mais il faudra toujours parvenir à un compromis entre ces exigences. Pour trancher la question, il convient alors de se demander si les preuves disponibles suggèrent que l'évaluation est suffisamment « adaptée à l'objectif visé » pour être utile. Par conséquent, il est recommandé de réfléchir à la validité en s'intéressant à la solidité des preuves, c'est-à-dire à l'« argument de validité » qui est avancé, plutôt que chercher une réponse immédiate par oui ou non. Cette idée est davantage développée dans la citation qui suit.

Si la déclaration de validité agit comme une promesse ou une garantie, qui donne aux utilisateurs le feu vert pour interpréter les résultats de l'évaluation et les utiliser aux fins indiquées, cela soulève un dernier point fondamental qui ne reçoit pas toujours l'attention qu'il mérite : sans formulation explicite de la prétention à la validité, les utilisateurs ne pourront pas interpréter ni utiliser les résultats de l'évaluation comme il se doit.

(Newton, 2007, p. 149 – 170)

Enfin, la validité n'est pas quelque chose qui est atteint ou non lors de la conception d'une évaluation, mais qui se développe en permanence au cours de son cycle de vie. De même, l'argument de validité n'est pas avancé au début du cycle : il est constamment nourri et précisé au cours du processus.

Qu'est-ce qui est valide : l'évaluation ou son utilisation ?

Réfléchissez à la situation suivante.

Un candidat a étudié la physique pendant 2 ans et a obtenu une note finale de 6. Il a passé l'examen en espagnol.

Cette note finale signifie-t-elle que le candidat parle couramment espagnol ? En poussant cet exemple jusqu'à l'absurde, cette note finale signifie-t-elle que le candidat est un bon cuisinier ?

Dans cet exemple, l'examen passé par le candidat a pu être un bon test de sa compréhension en physique, mais il a été utilisé comme une indication de sa compréhension de l'espagnol. L'évaluation est-elle donc valide ?

Un exemple plus subtil pourrait consister à se demander si une note finale particulière en mathématiques indique que le candidat est bon en arithmétique (soit un seul aspect du programme d'études). Cela introduit l'idée que ce n'est pas l'évaluation (ou le résultat) qui est valide, mais le but qu'elle poursuit. Newton (2012) a soutenu à ce propos que l'utilisation d'une procédure d'évaluation donnée dans un but

spécifique (c'est-à-dire pour prendre une décision particulière) était valide si son argument d'interprétation était suffisamment puissant.

Si c'est le but dans lequel l'évaluation est utilisée, et non l'évaluation elle-même, qui est valide, quelle responsabilité revient alors à l'IB dans la démonstration de la validité de ses examens ? Faut-il montrer la validité de chaque utilisation possible ? Est-il acceptable de définir un nombre limité d'utilisations pour lesquelles l'IB peut prétendre à la validité ? Les utilisations pour lesquelles l'IB ne prétend à aucune validité doivent-elles être explicitement énoncées ?

Une solution pratique doit être apportée à ce problème. Les évaluations sont conçues pour un certain nombre d'utilisations que l'IB pourra justifier en avançant un « argument de validité ». La conformité avec les autres utilisations courantes qui relèvent de la philosophie de l'IB sera également étudiée. Si l'analyse de ces deux catégories d'utilisation montre que les arguments sont suffisamment forts, alors la validité de l'évaluation sera indiquée de manière informelle.

Création d'un argument de validité

La validité n'est pas un concept objectif simple : il s'agit d'un juste milieu entre des questions opposées. Cela signifie que la validité ne peut pas être « prouvée », mais qu'il faut construire une argumentation convaincante qui montre pourquoi les choix opérés ont conduit à un cours et une évaluation pertinents.

Pour des raisons similaires, la preuve d'un argument de validité doit émaner naturellement du processus d'élaboration, de contrôle et d'enseignement d'un programme, et rendre compte des discussions et des décisions prises. Pour prendre un exemple dans le domaine de l'évaluation, les jugements à l'origine des questions posées dans une épreuve donnée devront prouver que le programme d'études est convenablement couvert.

Il est indispensable que l'argument de validité ait une structure appropriée couvrant tous les aspects considérés comme importants. À cet effet, l'IB a formulé une série de questions pour lesquelles des preuves sont réunies. Ces questions peuvent être consultées dans l'[annexe 2 « Feuille de route pour créer un argument de validité »](#).

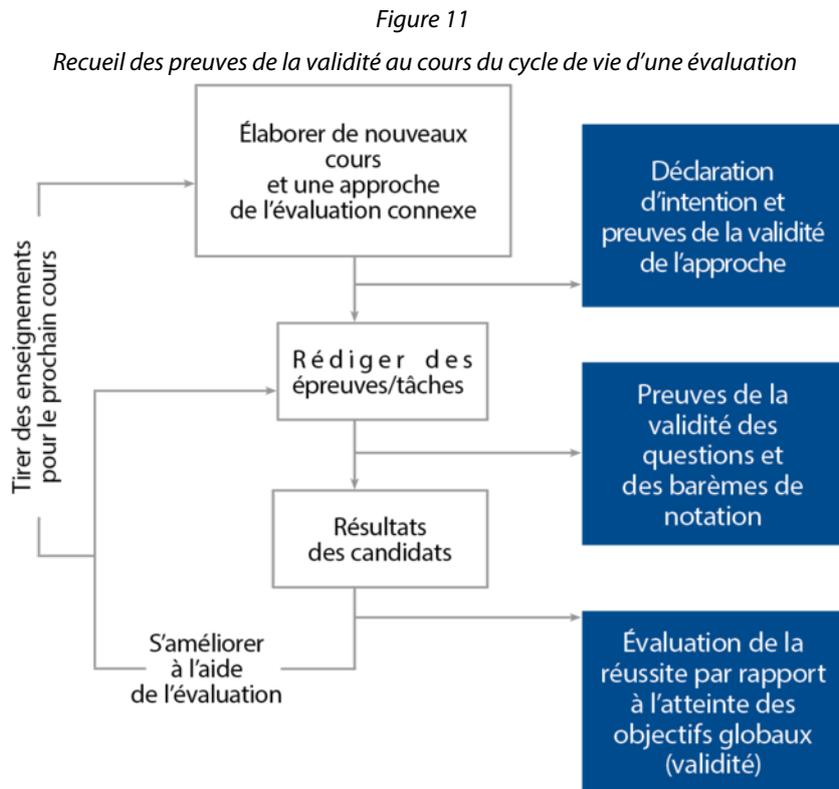
Maintien de la validité

La validité n'est pas quelque chose qui se décide ou se « prouve » au moment où un modèle d'évaluation est conçu. Elle se développe avec l'évaluation, au cours de son cycle de vie, et même au-delà, tant que les décisions prises sont basées sur son résultat.

L'IB rassemblera différentes preuves pour démontrer que l'évaluation est valide (c'est-à-dire l'argument de validité), à plusieurs moments du cycle de vie d'un cours.

Lorsque ce cours est élaboré ou révisé, les discussions sur le but du cours et la manière de l'évaluer pour atteindre ce but constitueront la partie centrale de l'argument de validité, et notamment l'équilibre entre la pertinence de construit et les autres aspects de la validité. Chaque session, la rédaction d'une nouvelle épreuve ou tâche d'évaluation fournira des preuves supplémentaires, notamment dans le domaine de l'équité, de la comparabilité et de la pertinence de construit. Les retours d'information des enseignants et les résultats des candidats permettront de juger dans quelle mesure l'évaluation parvient à atteindre son objectif, mais aussi à apporter des preuves de la fiabilité de la notation et de la comparabilité des notes finales. Ces informations seront ensuite utilisées pour orienter le développement de la prochaine série d'évaluations.

Enfin, les preuves rassemblées dans toutes les évaluations concernant un cours particulier, et dans tous les cours formant un programme donné, serviront de base aux décisions pour la prochaine révision du cours ou du programme. Ce processus est illustré par le schéma qui suit.



Même si toutes les preuves de l'argument de validité ne peuvent être recueillies qu'une fois l'évaluation passée par les candidats, chaque stade pose les bases de l'étape suivante du processus.

L'évaluation de la validité d'un test n'est pas une démarche statique et unique : il s'agit d'un processus continu.

(Sireci, 2007, p. 477)

En quoi l'ordinateur peut-il renforcer la validité ? Quels avantages l'évaluation électronique présente-t-elle en matière de validité ?

- Aujourd'hui, le monde n'utilise plus uniquement les supports papier : les ordinateurs ont pris une grande place dans tous les aspects de la vie.
- Les examens sur papier ne permettent pas l'utilisation d'images en mouvement ou une interaction intéressante avec les candidats.
- Grâce à la polyvalence des ordinateurs, les candidats peuvent apporter eux-mêmes les modifications visuelles et sonores nécessaires pour mieux prendre connaissance des questions, et éviter ainsi d'avoir à demander des épreuves spéciales plusieurs mois avant les examens.
- Contrairement aux idées reçues, l'évaluation électronique offre davantage de protection contre la mauvaise conduite et la mauvaise administration.

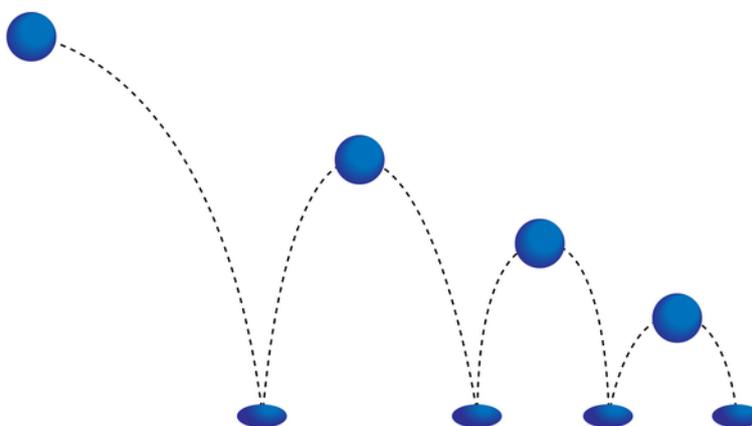
Il ne fait presque aucun doute que vous lisez ce document sur un écran, étant donné qu'il n'est disponible qu'au format électronique. Les courriels, les textos et les médias sociaux sont devenus des moyens de communication bien plus courants que les lettres, et dans bien des métiers, utiliser un ordinateur n'est plus

une pratique exceptionnelle, mais quotidienne. Si l'objectif est de proposer des évaluations authentiques, alors cela doit passer par une utilisation des ordinateurs.

Depuis toujours, l'évaluation est très limitée quant au spectre qu'elle peut couvrir. En règle générale, avec les examens écrits, les candidats doivent se contenter de réagir à un stimulus élémentaire ou de répondre à une question simple, sans avoir aucune possibilité de manipulation ni d'interaction. À son niveau le plus élémentaire, l'évaluation sur ordinateur permet d'inclure des éléments audio et vidéo en tant que stimuli. Les candidats activent alors eux-mêmes les fonctions « lecture » et « pause », et n'ont plus à attendre que le surveillant lance la lecture pour l'ensemble de la classe. Ce système permet également de remplacer une image et une longue description du mouvement d'un objet par un schéma animé.

Figure 12

La clarté d'une question d'évaluation électronique sur la trajectoire d'une balle rebondissante pourrait être améliorée grâce à l'utilisation d'une image en mouvement plutôt que d'une image statique, comme on le voit ici.



Avec l'évolution de l'évaluation électronique, il devient possible d'évaluer la manière dont un candidat traite un problème, réagit à de nouvelles informations ou développe une simulation. À terme, l'évaluation sur ordinateur devrait permettre de réagir aux pistes lancées par les candidats. Cela recréera les possibilités offertes par les évaluations organisées par les enseignants, comme les épreuves orales, en évitant les problèmes engendrés par le fait que les candidats ont différents enseignants.

Le principal atout de l'évaluation électronique est qu'elle permet d'évaluer ce que l'on souhaite véritablement évaluer, sans être limité par les capacités d'évaluation d'un examen sur papier.

La validité se trouve également renforcée par le fait qu'elle limite le biais, ce qui profite plus particulièrement aux candidats ayant des besoins en matière d'aménagement de la procédure d'évaluation. L'IB traite en effet régulièrement des demandes d'impression des épreuves d'examen dans des polices de caractères différentes, sur du papier d'une autre couleur, ou encore en braille.

L'évaluation sur ordinateur ne permet pas de supprimer ces obstacles, mais elle offre la possibilité aux candidats de choisir la taille de la police et la couleur qui sont les plus adaptées à leurs besoins. Et les avantages en matière d'accessibilité ne s'arrêtent pas là : les lecteurs d'écran constituent un atout de taille. Si certains candidats ne peuvent pas compter sur l'ordinateur pour répondre à leurs besoins en matière d'aménagement de la procédure d'évaluation, l'évaluation sur ordinateur est accessible à un éventail plus large de candidats, et ce, sans recourir à une aide supplémentaire, même si celle-ci reste envisageable si nécessaire.

Figure 13

Dans quels cas les évaluations électroniques pourraient-elles présenter un risque ?



Beaucoup craignent que l'évaluation électronique soit moins sûre que les examens sur papier. En réalité, les évaluations sur ordinateur sont à l'abri de certains risques auxquels sont exposés les examens sur papier, mais en courent d'autres.

L'un des principaux avantages de l'évaluation électronique est qu'elle peut être envoyée à travers le monde par voie électronique et rester « verrouillée » jusqu'à ce que le mot de passe soit communiqué aux candidats le jour de l'examen. Cet envoi est donc préférable à celui des épreuves sur papier, qui peuvent être lues par n'importe qui et imposent donc aux établissements de les conserver en lieu sûr pendant plusieurs jours avant la tenue de l'examen. Même si le système de sécurité de l'établissement était violé, les pirates auraient encore à pénétrer dans les évaluations électroniques, qui ont été spécialement conçues pour empêcher ce type d'attaque. Et si les pirates y parvenaient, ils se retrouveraient alors au même point qu'avec des examens sur papier volés, mais ils disposeraient probablement de moins de temps pour en faire usage, car les évaluations électroniques sont transmises aux établissements à une date plus proche des examens que les épreuves sur papier.

Il faut admettre qu'un problème est susceptible de se présenter le jour de l'examen, étant donné que les candidats ont accès à un ordinateur qui, en théorie, peut leur permettre d'accéder plus facilement à des notes non autorisées que les méthodes traditionnelles de dissimulation des notes dans la salle d'examen. Les évaluations électroniques de l'IB sont toutefois conçues pour fonctionner en mode « kiosque », ce qui empêche toute utilisation simultanée d'un autre programme. De plus, il est possible d'enregistrer la manière dont les candidats répondent aux questions, en plus de leurs réponses définitives, et donc de détecter tout comportement assimilable à une tricherie.

En bref, l'évaluation électronique pose de nouvelles difficultés concernant la sécurité des examens, mais elle supprime aussi certaines occasions de tricher qui existaient par le passé.

Éléments de la chaîne de validité

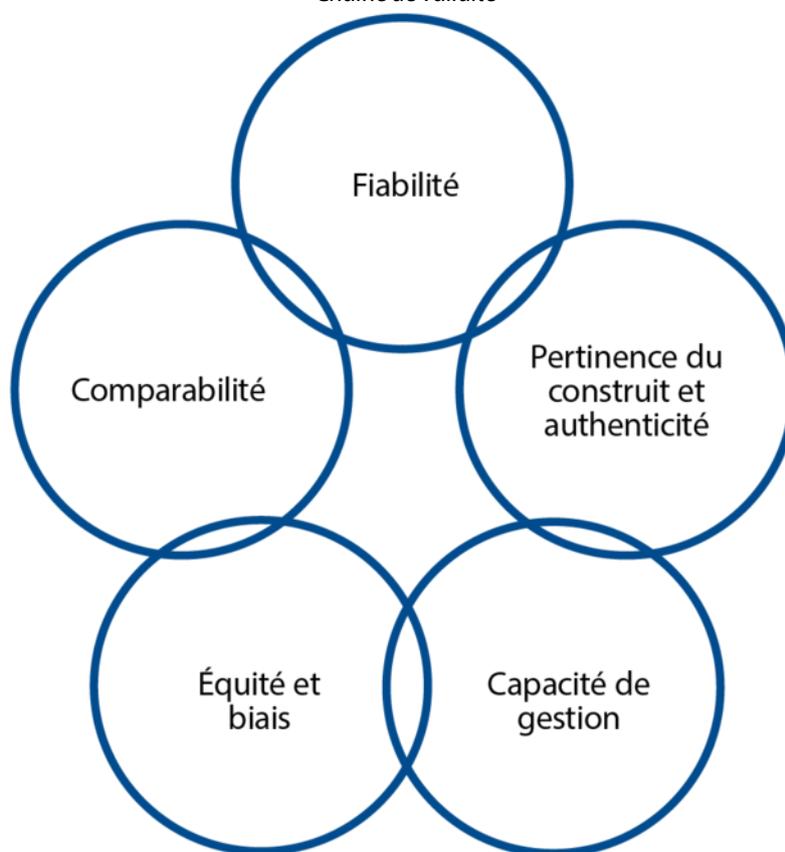
- La validité comporte de nombreux aspects. En imaginant qu'ils forment ensemble une chaîne, si un seul maillon est cassé, alors c'est toute l'évaluation qui est détruite.
- L'IB se concentre généralement sur cinq éléments de la validité : la fiabilité, la pertinence de construit, la capacité de gestion, l'équité et la comparabilité.
- Ces cinq éléments entrent souvent en contradiction les uns avec les autres. Pour savoir si une évaluation est valide, il faut s'intéresser au but premier de l'évaluation afin de déterminer l'importance relative de chaque aspect.
- L'IB s'attache avant tout à concevoir des programmes, des cours et des évaluations dont le construit est pertinent.
- Dans le domaine de l'évaluation, cela signifie que l'accent est placé sur l'élaboration de tâches et de questions pertinentes, qui mesurent les compétences de pensée de haut niveau, plutôt que sur la création d'examens garantissant une notation fiable.

La validité (au même titre que la fiabilité) est largement considérée comme une caractéristique essentielle de tout système d'évaluation, et les enjeux sont d'autant plus élevés que le résultat revêt une grande importance pour le candidat ou l'enseignant. Les caractéristiques sont en réalité multidimensionnelles, et les types de validité et de fiabilité sont nombreux.

L'idée de la chaîne montre la complexité de la validité (Crooks, Kane et Cohen, 1996). Chacun des cinq maillons de la chaîne revêt une importance en soi, mais aucun ne suffit à lui seul à rendre l'évaluation valide (c'est-à-dire adaptée à l'objectif visé). Ainsi, une évaluation peut être très fiable, mais désavantager systématiquement un groupe donné. Il est également possible qu'une tâche atteigne précisément l'objectif visé, mais qu'elle mette également à rude épreuve l'endurance des candidats et les ressources de l'établissement en raison de sa durée et de ses exigences.

Pour qu'une évaluation soit valide, tous les maillons de la chaîne doivent coexister.

Figure 14
Chaîne de validité



Les sections suivantes donnent plus de précisions sur chacun des éléments de cette chaîne. Même si tous sont nécessaires pour parvenir à la validité, il convient de souligner qu'il existe une certaine tension entre eux. Par exemple, si un examen doit aborder tous les aspects du programme d'études, cela peut facilement mener à des tests trop longs pour les candidats. Dans le même ordre d'idée, si l'on souhaite élaborer un test en tenant compte du contexte propre à chaque culture ou pays (équité), il faut alors se demander si les candidats passeront bien tous le même test (comparabilité).

Trouver un équilibre entre les différents aspects de la validité

Lors de la conception d'une évaluation, il est important de trouver un équilibre entre les priorités contradictoires, généralement en tenant compte du but visé par l'évaluation. Aucune évaluation ne peut satisfaire à la norme la plus élevée de tous les éléments ; c'est pourquoi il faut parvenir à un compromis. Alors que certains de ces éléments sont fixés lors de la conception de l'évaluation, d'autres (notamment la fiabilité et l'équité) évoluent avec l'utilisation de l'évaluation, la notation et l'attribution des notes finales.

Figure 15

Trouver un équilibre entre les priorités contradictoires de plusieurs aspects de la validité : l'équité et la capacité de gestion

		
<p>Mauvais équilibre La charge pesant sur les candidats est excessive (capacité de gestion)</p>	<p>Bon équilibre Les candidats sont évalués de manière raisonnable</p>	<p>Mauvais équilibre Il y a une trop grande part de chance dans le choix du thème évalué (équité)</p>
<p>L'évaluation est composée de 50 tâches (de 20 minutes chacune) requérant une réponse développée et traitant tous les aspects du cours. Cela représente plus de 16 heures d'évaluation.</p>	<p>L'évaluation comporte 20 questions à réponse brève (de 2 minutes chacune), 5 tâches plus longues (de 10 minutes chacune) et 2 tâches approfondies (de 30 minutes chacune). Chaque question porte sur un aspect différent du cours. L'évaluation dure deux heures et demie en tout, et porte sur un peu plus de la moitié des thèmes du cours.</p>	<p>L'évaluation consiste en une tâche approfondie de 30 minutes, qui mesure à la fois les connaissances et la compréhension d'un aspect du cours (sur les 50 disponibles). La note finale du candidat repose exclusivement sur une tâche.</p>

Les sections suivantes montreront que, malgré des définitions distinctes pour les cinq éléments de la validité, il existe des recoupements importants concernant leur intervention et leur gestion. Ces éléments constituent également des aspects du concept de validité au sens large.

Enfin, l'IB accorde une priorité absolue à la validité de construit, c'est-à-dire que ses évaluations mesurent les caractéristiques et les capacités qu'elles sont censées mesurer. Il convient toutefois de rappeler que ce sens des priorités ne peut se faire totalement au détriment des autres aspects de la chaîne de validité.

Fiabilité

La fiabilité se définit comme « la mesure dans laquelle un candidat pourrait obtenir le même résultat à un test si la procédure employée était répétée ».

Comme il est indiqué ci-après, cela n'est pas nécessairement la même chose qu'un candidat obtenant le « bon résultat ».

Dans leur introduction au concept de fiabilité, Winkley et Cresswell (dans l'ouvrage *Ofqual's Reliability Compendium* publié en 2011) dressent une liste (non exhaustive) des sources d'un éventuel manque de fiabilité.

1. La fiabilité entre les correcteurs : un examinateur pourrait être plus ou moins indulgent qu'un autre pour certaines questions (voire montrer lui-même un degré d'indulgence différent d'un jour à l'autre).
2. La variabilité des résultats d'un candidat : les résultats d'un candidat à un examen peuvent varier légèrement d'un jour à l'autre, notamment si les conditions d'examen sont différentes (matin ou après-midi, personne faisant passer le test, qualité du sommeil la veille de l'examen, bruit de la tondeuse pendant l'examen, mal de tête, etc.).

3. La différence entre les épreuves d'examen : les questions sont différentes d'une épreuve d'examen à l'autre et risquent de tester des aspects différents de la compréhension du candidat (les tests portent généralement sur un échantillonnage de thèmes du programme d'études en raison d'un manque de temps pour tester la totalité, et les candidats peuvent choisir de ne pas réviser tous les thèmes).
4. La comparabilité des résultats d'une année à l'autre.
5. Les différences entre les modalités des examens : il peut être difficile de garantir la comparabilité dans le temps lorsque des changements sont apportés aux modalités des examens et aux programmes.
6. Les différents types d'activités d'évaluation : de nombreuses évaluations sont composées de plusieurs types d'activités, et ces différentes méthodes d'évaluation peuvent engendrer plusieurs problèmes de fiabilité.
7. Les différents types de questions : les candidats pourront obtenir des résultats différents en fonction des types de questions qu'ils doivent traiter.

Dans la pratique, l'IB part généralement du principe que le candidat passe l'évaluation une seule fois et cherche l'homogénéité du résultat du processus, d'où une priorité donnée aux points 1, 3, 6 et 7 de la liste.

En tant qu'organisme décernant des titres, l'IB prend des mesures pour accroître le niveau de fiabilité de ses évaluations. La fiabilité de la notation occupe une place centrale dans cette démarche, et la normalisation, le modèle d'assurance de la qualité et la révision de notation visent à garantir que les examinateurs procèdent à la notation selon une norme cohérente.

Fiabilité de la notation

Le rapide exercice ci-dessous illustre le concept de fiabilité de la notation et son fonctionnement dans la pratique.

Cinq extraits de réponses authentiques de candidats à une évaluation en acquisition de langues de l'IB sont présentés. Cliquez sur ces réponses pour visualiser une par une la copie de chaque candidat.

Utilisez le barème de notation ci-dessous pour juger de la qualité de langue démontrée par les candidats et consignez la note sur 10 que vous attribueriez à chaque travail.

Après avoir « noté » chaque réponse, consultez le retour d'information sur la qualité de votre notation. Ce retour d'information est fourni plus bas, après les copies des candidats.

Barème de notation

Avec quel degré d'efficacité et de précision l'élève utilise-t-il la langue ?

Points	Descripteurs de niveaux
1 – 2	<p>La maîtrise de la langue est généralement insuffisante.</p> <p>Un vocabulaire très limité est utilisé, avec de nombreuses erreurs de base.</p> <p>Des structures grammaticales simples qui font rarement preuve de clarté.</p>
3 – 4	<p>La maîtrise de la langue est limitée et généralement inefficace.</p> <p>Un vocabulaire limité est utilisé, avec de nombreuses erreurs de base.</p> <p>Des structures grammaticales simples qui font parfois preuve de clarté.</p>
5 – 6	<p>La maîtrise de la langue est généralement satisfaisante, en dépit de nombreuses inexactitudes.</p> <p>Un vocabulaire assez limité est utilisé, avec de nombreuses erreurs.</p> <p>Des structures grammaticales simples qui font généralement preuve de clarté.</p>
7 – 8	<p>La maîtrise de la langue est efficace, en dépit de quelques inexactitudes.</p> <p>Le vocabulaire est adapté à la tâche et varié, avec quelques erreurs.</p> <p>Des structures grammaticales simples qui font preuve de clarté.</p>

Points	Descripteurs de niveaux
9 – 10	<p>La maîtrise de la langue est satisfaisante et efficace.</p> <p>Le vocabulaire est adapté à la tâche et varié, avec peu d'erreurs importantes.</p> <p>Diverses structures grammaticales complexes sont utilisées efficacement et font preuve de clarté.</p>

Visualisez chaque copie une par une et attribuez une note à chacune en vous aidant du barème de notation.

Figure 16

Exercice sur la fiabilité de la notation

Candidat A

Dans notre monde tellement dominé par les réseaux sociaux, il me semble que les gens ne savent plus distinguer le vrai et le faux qui s'affiche devant leurs yeux sur l'écran de leurs smartphones. Malheureusement, la superficialité domine, pas l'information... c'est triste.

Aujourd'hui au collège, j'étais en cours avec Pierre et Louis de l'article sur les extraterrestres qui est devenu "viral" sur Facebook l'autre jour. C'était un article qui essaye de convaincre le public que les aliens ont fait contact avec nous et qu'ils aiment bien le climat de l'Alaska... grosse blague.

Candidat B

La réflexion « Dans le réseaux sociaux, c'est la superficialité qui domine, pas l'informations relié au thème communication et media, car sa concerne le reseaux sociaux, et relié aussi à question mondiale car le information conserve tout le monde.

Pour commencer, je pense que le réseaux sociale aide beaucoup pour faire connaître tout qui entraîne de passer au temps réel, sa aide a donner tout le information, le danger, l'histoire, etc. Par exemple si quelqu'un qui vit pas en Amérique voulez savoir comme la votation est passé, le réseaux sociaux a permettre a tout le jeune de suivre sur l'internet.

Candidat C

C'est vrai qu'avec Instagram, par exemple, il y a un grand aspect de superficialité, où l'on veut tous créer l'image d'un vie 'parfaite'. C'est-à-dire seulement publier des photos de nous sur son site. Par exemple, pendant les attentats à Paris et en Florides, il était possible de se marquer 'safe' (en sécurité), et ceci notifié les amis. Du coup, Facebook a utilisé des paramètres pour informer les amis et les familles que la personne qui était près du danger était en sécurité.

Candidat D

En revanche, quand on sait à utiliser les réseaux sociaux responsable dans une manière responsable, l'un peu dire que l'individu pourra se sauver d'être un accro à les réseaux sociaux par l'utilisation ^{en modération} modérée. Le savant va réfléchir l'impact des images ou informations il choisit à partager. Parfois les gens veillent le sensationnel dans leur réseau sociaux car c'est intéressant pour consumer, alors, des sites, comme «wisecrack» nous fournissent avec des informations culturelle et actuelle en forme comédie mais équilibré avec les faits réelle.

Candidat E

la réflexion « Dans le réseaux sociaux, c'est la superficialité domine, pas l'informations reliées au thème communication et media, car sa conserve le reseaux sociaux, et relie aussi à question mondiale car le information conserve tout le monde.

Pour commencer, je pense que le réseaux sociaux aide beaucoup, pour faire connaître toute qui entraîne de passer au temp reale, sa aide a donner tout le informations, le danger, l'histoire, etc. Par exemple si quelqu'un qui vivre pas en Amérique voulez savoir comme la votation est passé, le réseaux sociaux à permettre a tout le jeune de suivre sur l'internet.

↳

Vous devriez maintenant avoir terminé votre « notation » de chaque copie et avoir attribué une note à chaque candidat.

Lisez maintenant le retour d'information sur votre notation.

Retour d'information

Avez-vous remarqué que le travail du candidat B et du candidat E était le même, mais avec une écriture différente ?

Avez-vous donné la même note au candidat B et au candidat E ?

Vous avez peut-être aussi hésité entre les deux notes pour une bande de notation donnée. Dans quelle mesure êtes-vous sûr(e) que vous prendriez la même décision si vous deviez noter de nouveau ces travaux ? Il s'agit ici d'attirer votre attention sur le point suivant : ce qui est en cause n'est pas la qualité du travail de l'élève, mais la façon dont la même personne chargée de noter un travail peut formuler des jugements légèrement différents à chaque fois qu'elle note ce travail. Globalement, votre opinion sur la qualité du travail restera la même à chaque fois, mais la note exacte que vous attribuerez pourra varier légèrement.

Refaites cet exercice dans quelques jours en notant de nouveau chaque copie, et comparez cette deuxième notation à la première. Ne tentez pas de vous souvenir des notes que vous aviez données initialement.

Résultats cohérents et bons résultats sont deux choses différentes

Il est important de comprendre que l'objectif consistant à obtenir un niveau de fiabilité élevé permet de garantir que les candidats auront la même note (juste) quel que soit l'examineur passant leur travail en revue, et non la « bonne » note. Pour déterminer la qualité d'un travail, l'enseignant s'appuie sur son jugement professionnel, et deux enseignants auront souvent du mal à tomber d'accord sur la note à attribuer. La fiabilité intervient lorsque tous deux émettent le même jugement (identique à celui de l'examineur superviseur).

Cela présente donc des difficultés particulières lors du traitement des réclamations concernant les résultats. Dans ce cas, l'examineur doit veiller à attribuer la même note que celle qu'il aurait attribuée si la copie avait été la première qu'il notait, et s'assurer que cette note n'est pas influencée (positivement ou négativement) par de nouvelles informations, comme les seuils d'attribution des notes finales ou les conséquences pour le candidat concerné.

La mauvaise compréhension de la fiabilité de l'évaluation dans des contextes autres que l'éducation a fait l'objet de nombreuses publications, mais les débats sur les résultats aux examens étant de plus en plus courants, une importance particulière doit être accordée à ce thème.

Comme la littérature semble l'indiquer, les participants ont estimé que la fiabilité de l'évaluation et plus précisément l'inexactitude de la mesure étaient des concepts difficiles à saisir.

(Chamberlain, 2010, p. 3)

Je me dis parfois que les examens ne permettent pas de jauger l'aptitude d'une personne à sa juste valeur. Ils reflètent davantage comment la personne allait à ce moment précis, sa faculté à étudier, etc. Ce n'est pas parce que quelqu'un obtient d'assez mauvais résultats aux examens qu'il ne pourrait exceller par la suite et compter parmi les meilleurs de son domaine. (Propos d'un employé du secteur de la santé)

(Chamberlain, 2010, p. 27)

Environ 63 % des enseignants et des élèves ont choisi « Aucun niveau d'erreur ne doit être acceptable ; même le fait qu'un seul candidat obtienne une note finale incorrecte est totalement inacceptable », mais plus de 50 % d'entre eux ont également choisi « Il existe une différence entre une erreur évitable – comme une coquille dans une épreuve – et quelque chose d'inévitable comme un manque de cohérence entre deux correcteurs », ce qui suggère une tolérance envers l'erreur. Cette incohérence peut illustrer la faiblesse du lien entre les connaissances sur la fiabilité et les attitudes envers le manque de fiabilité.

(He, Opposs et Boyle, 2010, p. 27)

Pertinence de construit et authenticité

Avec quelle précision la chose censée être mesurée est-elle véritablement mesurée ? Les expressions « pertinence de construit » et « validité de construit » sont parfois utilisées indifféremment pour se référer au même concept. Cependant, pour éviter toute confusion avec la notion générale de validité, seule l'expression « pertinence de construit » sera retenue ici.

L'idée d'évaluation authentique est étroitement liée à la notion de pertinence de construit. Cela signifie que les modalités du test correspondent aux situations dans lesquelles le candidat pourrait s'attendre à rencontrer le problème dans la vie réelle. Les évaluations qui isolent les tâches de leur contexte, qui simplifient les choses à outrance ou qui sont manifestement artificielles, sont des exemples d'évaluations non authentiques.

Il n'est pas compliqué de trouver des exemples d'évaluations mal conçues, avec un faible degré de pertinence de construit (par exemple, évaluer la capacité à rédiger une lettre en utilisant une épreuve orale), mais, souvent, même la méthode reconnue en matière d'élaboration des tâches ne permet pas d'évaluer véritablement l'objet visé. Prenez par exemple la conception de certains examens de littérature traditionnels. Les candidats ont étudié les procédés littéraires qui sont utilisés dans un texte donné, puis l'évaluation leur demande de rédiger une dissertation en rapport avec l'un de ces procédés. Si les candidats

doivent simplement se souvenir de ce que l'enseignant leur a appris, alors montrent-ils réellement une compréhension de la littérature ?

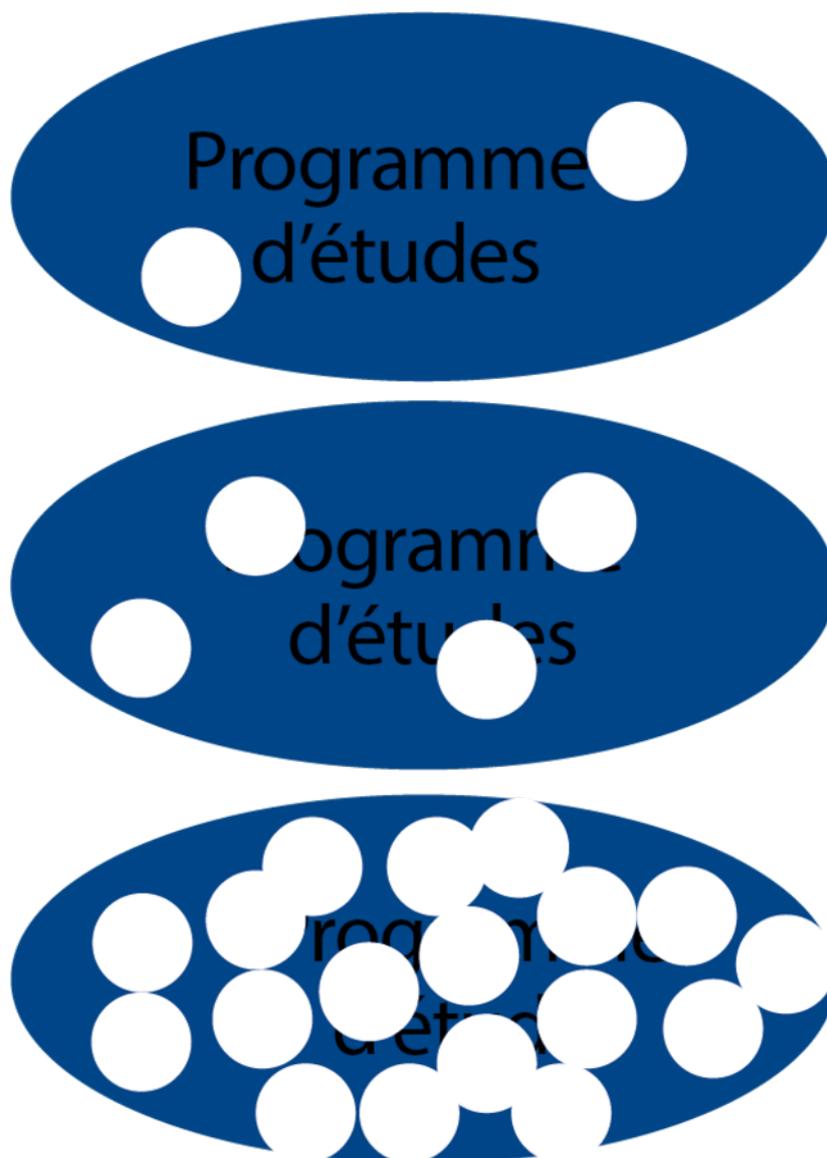
Bien comprendre ce qu'une tâche d'évaluation donnée souhaite permettre au candidat de montrer, et analyser avec soin dans quelle mesure elle parvient à le faire, constitue la meilleure pratique pour s'assurer de l'adéquation du degré de pertinence de construit. Il convient notamment de réfléchir aux autres compétences dont les candidats ont besoin pour réaliser la tâche. Ainsi, les tâches ouvertes, comme les dissertations ou les projets, nécessitent très souvent des compétences élevées en matière de rédaction pour que les candidats puissent montrer les compétences de recherche ou d'analyse visées par la tâche. Cette idée est reprise dans le concept de conception universelle de l'évaluation, qui s'attache à réduire les obstacles pour les candidats lorsque les difficultés n'ont aucun rapport avec ce que l'évaluation doit évaluer (Dolan, Burling, Harms, Strain-Seymour, Way et Rose, 2013).

Sur le plan historique, l'approche de la mesure de la pertinence (ou validité) de construit a très largement été inspirée de la psychométrie. Il en a résulté un cercle vicieux, dans lequel ce que le test évaluait (autrement dit, le construit) était défini par le test lui-même. Si toutes les questions du test présentaient un résultat cohérent, alors cela était perçu comme une preuve de la pertinence de construit du test. Cependant, si le résultat d'une question n'était pas aligné sur celui des autres questions, cela indiquait que la pertinence de construit n'était pas au rendez-vous et, par conséquent, la question était écartée. Les approches modernes s'éloignent généralement de cette définition purement statistique et autoréférentielle de la pertinence de construit.

La pertinence de construit est étroitement liée au programme d'études, comme il est indiqué dans la figure 2 intitulée « Lien entre les objectifs pédagogiques, les objectifs d'apprentissage et la conception de l'évaluation ». Très souvent, l'évaluation interroge l'élève uniquement sur une partie limitée du programme d'études, et la question consistant à se demander si ce qui est traité est suffisant relève de la pertinence de construit. De même, le choix des types de questions qui seront posées constitue un élément important de cette conception, car chaque type de question ou de tâche possède des points forts et des points faibles différents pour ce qui est du type de construit qu'il peut évaluer.

Figure 17

Chaque examen (cercle blanc) porte sur une partie du programme d'études.



Capacité de gestion

Contrairement aux deux éléments précédents, la capacité de gestion a fait l'objet de peu de recherches, et il n'existe pas de définition ni d'approche unique de sa mesure. En termes généraux, cela désigne l'effort requis pour passer une évaluation, et il est raisonnable d'aborder la capacité de gestion du point de vue du candidat, de l'établissement scolaire et même de l'IB.

Pour les candidats, la capacité de gestion fait souvent référence à la durée de l'évaluation. Ainsi, un examen de 8 heures est considéré comme une exigence excessive pour un élève de 18 ans. Il faut également tenir compte de l'heure à laquelle a lieu l'évaluation. Par exemple, il peut sembler raisonnable de prévoir une série de quatre examens, chacun d'une durée d'une heure, mais si ces examens ont lieu le même jour et qu'ils ne sont séparés que d'une courte pause, alors la charge peut être lourde. Le contexte de l'IB implique de faire particulièrement attention à la charge induite par les différentes évaluations de l'IB que les

candidats effectueront dans le cadre de leurs études, ce qui ramène à l'idée qu'il ne faut pas considérer uniquement la validité de chaque matière, mais celle des programmes.

Du point de vue de l'établissement scolaire, la capacité de gestion peut également inclure les exigences en matière de fourniture de matériel pour l'évaluation. Ainsi, une formation professionnelle en ingénierie pourrait exiger que chaque candidat assemble un moteur. Dans le contexte de l'IB, l'exigence consistant à pouvoir passer l'évaluation sur un ordinateur ou tout autre appareil similaire imposerait certainement de se demander si une telle évaluation est gérable.

Un autre aspect de la capacité de gestion pour les établissements scolaires concerne la manière dont les travaux des candidats sont transmis à l'IB en vue de la notation. L'obligation d'enregistrer ou de filmer un exposé impose davantage de contraintes à l'établissement que l'envoi postal d'un travail écrit.

Enfin, la capacité de gestion concerne également l'IB, en raison de la quantité de travaux de candidats passés en revue. Ainsi, assister à une production théâtrale de trois heures peut constituer la meilleure preuve de l'aptitude d'un candidat, mais cette pratique n'est pas envisageable si la tâche doit être évaluée en externe. Pour prendre un autre exemple, il ne serait pas raisonnable de prévoir un large éventail de questions facultatives si l'établissement d'une norme commune entre ces questions implique un travail à la fois long et minutieux.

Il est souvent difficile de concilier la capacité de gestion, la fiabilité et la pertinence de construit. Toute augmentation de la quantité d'évaluation (durée) fournira davantage de preuves de la compréhension qu'a le candidat du programme d'études et augmentera la probabilité d'un lissage entre la générosité et la sévérité des décisions en matière de notation. Cependant, une telle augmentation conduirait également à une évaluation mesurant l'aptitude des candidats à maintenir leur rendement pendant un examen long au lieu des objectifs de l'évaluation.

L'IB procède à des contrôles stricts de la capacité de gestion de l'évaluation, notamment pour ce qui est de la durée totale de l'évaluation dans chaque cours.

Équité et biais

Un test est biaisé s'il confère un avantage à une personne sans que cela soit le but de l'évaluation. Pour replacer les choses dans leur contexte, quelques exemples s'imposent.

- Un examen d'histoire entièrement rédigé en latin.
- Un examen de mathématiques dans lequel toutes les questions se rapportent au marquage de points dans un match de cricket.
- Un examen d'arts (peinture) pour lequel le chevalet est fixé à deux mètres du sol.

Dans chaque situation, les tâches (questions) elles-mêmes peuvent constituer un test raisonnable, mais certains candidats (ceux de petite taille pour le dernier exemple) seront considérablement défavorisés. Dans la réalité, les exemples de biais sont généralement plus subtils que ceux donnés précédemment, mais, à moins d'y faire véritablement attention, le biais peut provoquer des différences importantes en ce qui concerne les résultats des candidats. Placer les questions dans un contexte se révèle particulièrement difficile, car les situations familières à certains candidats peuvent être inconnues d'autres candidats, ce qui se vérifie d'autant plus avec la dimension internationale de l'IB.

Le biais peut être défini comme une différence dans le résultat d'un processus d'évaluation qui n'est pas due à une réelle différence dans l'aptitude ou l'accomplissement mesuré. Le biais peut venir des conditions de l'évaluation, de la notation de l'évaluation (dans ce cas, le biais devient un problème de fiabilité de la notation) ou directement des questions ou des tâches d'évaluation.

Biais venant des conditions de l'évaluation

Les exemples donnés au début de cette section comportent un biais au concernant les conditions, parfois improbables comme le chevalet placé trop haut pour certains candidats, mais de nombreuses conditions sont susceptibles d'engendrer un biais pour une évaluation, et plus précisément pour un examen. Beaucoup concernent l'horaire de l'examen, qui est prévu par exemple alors que les températures ou le niveau de pollution sont extrêmes dans certaines parties du monde, ou à un moment où certains candidats

ne peuvent pas se concentrer ou se préparer convenablement. Cela pose un défi majeur à l'IB, étant donné que les exigences contradictoires de certains pays impliquent qu'aucune date n'est idéale pour au moins quelques établissements.

L'agencement de la salle peut lui aussi constituer un biais fréquent. Prenez le cas de deux candidats, le premier étant assis à une place directement exposée au soleil et, le second, à l'ombre. Le manque de cohérence dans l'application du règlement de l'examen peut fournir un autre exemple. Ainsi, alors qu'un candidat est scrupuleusement soumis aux délais impartis, un autre bénéficie d'une certaine souplesse. Pour pallier cette situation, l'IB s'appuie sur des règles claires et cohérentes rassemblées dans le document intitulé *Procédures d'évaluation* spécifique à chaque programme, et considère tout manquement à ces règles comme une mauvaise administration.

Biais venant de la notation

Il est important de commencer cette section en soulignant que le biais venant de la notation est bien plus inconscient que délibéré. Par nature, l'esprit humain utilise des raccourcis pour faciliter la prise de décisions, et il en résulte souvent des biais inconscients. Cela ne signifie pas que l'IB n'a pas le devoir de limiter ces biais, mais simplement qu'aucune faute ne lui est associée.

Lorsque le biais se manifeste lors de la notation, il peut être dû à un certain nombre de raisons, telles que l'attitude du correcteur à l'égard de la netteté de l'écriture du candidat (par exemple, Hughes *et al.*, 1983), un traitement préférentiel en fonction du sexe du candidat (lorsque celui-ci est connu du correcteur ou lorsqu'il le devine) et une attention excessive à des facteurs comme le formatage, la ponctuation et l'orthographe, qui ne sont pas très pertinents dans certains contextes d'évaluation. Pour résoudre ces problèmes, il suffit de former les correcteurs et de vérifier leur travail.

Le biais inconscient fondé sur des facteurs tels que le sexe, la nationalité ou l'établissement est également bien documenté. Pour le minimiser, l'IB cherche à rendre tous les travaux des candidats anonymes avant qu'ils soient notés.

Un autre domaine de biais bien étudié est connu sous le nom d'« effet de halo ». Cet effet peut être observé lorsqu'un examinateur développe une opinion positive sur un candidat en raison de la bonne qualité de ses premières réponses et que le bénéfice du doute devient disproportionné pour les questions suivantes.

Biais associé aux questions de l'évaluation

Le biais ne s'applique pas si la différence se rapporte à l'objet visé par la mesure. Par exemple, la différence de taille entre les hommes et les femmes n'est pas le résultat d'un biais dans la façon de mesurer la taille. Même si elle n'est pas biaisée, l'évaluation peut être discriminatoire.

Le biais provenant des tâches d'évaluation elles-mêmes est le problème le plus important. Lors de l'élaboration des tests psychométriques, tout élément de question pour lequel une réponse peu habituelle est donnée durant le test préliminaire, ou pour lequel les différents sous-groupes de candidats donnent une réponse très différente (« fonctionnement différentiel des éléments »), est considéré comme biaisé, et il est supprimé du test. Les sous-groupes de candidats peuvent être formés en fonction du sexe, de l'origine ethnique, de la classe sociale ou de la compétence linguistique des candidats. En fait, ils peuvent être formés en fonction de n'importe quelle caractéristique n'ayant aucun rapport avec le construit évalué.

Les déclarations de partialité pour ou contre certains sous-groupes de candidats ne sont pas toujours facilement justifiables. Durant les premières années du développement des tests d'intelligence, les éléments de question qui donnaient lieu à des réponses très différentes de la part des deux sexes étaient exclus. On se basait sur le principe selon lequel il ne devrait y avoir aucune différence d'intelligence entre hommes et femmes, et que, de ce fait, tout élément qui révélait une telle différence mesurait une chose non pertinente. Un tel point de vue est au moins discutable, et divers auteurs ont expliqué que ces différences d'intelligence entre les divers groupes de personnes étaient liées à des facteurs biologiques, environnementaux ou socioéconomiques, ainsi qu'à la nature des tests eux-mêmes.

Lors du développement des tests d'intelligence, on s'est davantage préoccupé de la fiabilité des mesures que de la nature de ce qui était en fait mesuré, laquelle a été travaillée pour satisfaire aux besoins de haute fiabilité. Il peut y avoir des aspects importants d'un construit qui sont légitimement liés à certaines caractéristiques des groupes composant l'ensemble de la population scolaire. Les implications de telles

différences pour l'enseignement et l'apprentissage sont considérables. L'une des méthodes préférées est de concevoir des tests contenant un nombre équilibré d'éléments de question donnant lieu à des résultats différents de la part des différents sous-groupes d'élèves, et ce, de façon à ce que, de manière globale, aucun sous-groupe ne soit désavantagé. Quelle que soit la méthode adoptée, elle peut engendrer d'énormes difficultés lors de la conception du test.

Au moment de décider si une question donnée ou un certain élément sont biaisés, il faut soigneusement établir de quelle façon la tâche en question peut être reliée de manière explicite au construit sous-jacent et déterminer les facteurs de biais possibles. Baser cette décision sur des considérations purement statistiques risque de faire tomber dans le piège de la confusion entre une évaluation biaisée et une évaluation dont l'objet établit une discrimination entre les groupes. Goldstein (1996) et Humphreys (1986) ont suggéré qu'il était utile de faire une distinction entre la « différence », qui est un fait déterminé de manière objective, et le « biais », qui est un jugement porté sur la pertinence de la différence. Black (1999) propose les six possibilités suivantes, qui sont les cas les plus communs où une question peut être injuste à cause de son impact sur les différents élèves.

- Le contexte dans lequel la question est posée (par exemple, les références culturelles américaines favorisent les élèves vivant aux États-Unis et peuvent défavoriser les élèves vivant ailleurs).
- Les sujets de dissertation portant sur les relations humaines favorisent les groupes de la société qui encouragent les émotions.
- Les questions à choix multiple peuvent avantager les garçons.
- Les composantes d'évaluation qui prennent la forme de travaux réalisés dans le cadre du cours ou de projets peuvent avantager les filles.
- Une question utilisant le langage ou les conventions d'une certaine classe sociale favorise les élèves issus de cette classe sociale.
- Certaines questions ne sont compréhensibles que dans certaines cultures. C'est notamment le cas d'un sujet d'examen portant sur le fait que les personnes âgées vivent seules, car ce concept est complètement étranger à certaines cultures, ou d'une question portant sur le rôle typique de la femme ou de l'homme dans une certaine culture, car elle peut paraître très déplacée dans une autre.

En ce qui concerne le biais fondé sur le sexe, il existe des preuves de ce biais, notamment aux États-Unis, mais il est impossible pour l'heure de déterminer avec certitude quels sont les aspects du format qui conduisent à ce bilan.

Au moment de concevoir les évaluations, la façon dont il faut tenir compte de ces différences n'apparaît pas toujours clairement. Les évaluations doivent être conçues de façon à ce que l'incidence globale du biais soit réduite. Pour ce faire, un éventail de tâches et de types de questions sera utilisé. Tout stéréotype lié à la culture ou au sexe (qu'il soit explicite ou non) doit être évité. Le contenu de chaque question doit être minutieusement examiné afin d'éviter les catégories les plus évidentes de questions injustes. De plus, les tests préliminaires, qui consistent à donner ces questions aux échantillons d'élèves des différents sous-groupes de la population scolaire, peuvent révéler des cas cachés. Toutefois, si l'on exclut tous les types de questions biaisées et tous les scénarios possibles, les concepteurs d'évaluations et les rédacteurs de sujets d'examen se retrouvent devant un choix limité, et les contraintes qui en résultent auront un impact négatif sur la validité de l'évaluation. À part éviter les pièges évidents et superflus, la solution la plus raisonnable semble être d'adopter une approche équilibrée lors de la conception de l'évaluation, en utilisant pour cela divers types de tâches d'évaluation et de présentations.

Des inquiétudes portent également sur le fait de savoir combien de sous-groupes différents d'élèves d'une même population scolaire doivent être pris en considération. Faut-il tenir compte des candidats ayant des styles d'apprentissage différents ou de ceux ayant un tempérament qui ne leur permet pas de passer des tests ou examens officiels ? Comme Hieronymus et Hoover (1986) l'ont indiqué, si les différences en matière d'intérêts et de motivation sont considérées comme des facteurs de biais, il est alors possible de considérer que toutes les tâches ou méthodes d'évaluation comportent un certain biais. Par exemple, certains extraits utilisés pour les examens de langue intéresseront sûrement plus certains candidats que d'autres. En fin de compte, la préoccupation concernant les possibilités de biais dans les différents types de questions et de contextes revient à une question de sociopolitique. L'équité dans le contexte de l'évaluation, qui comprend

notamment le soin d'éviter le biais, est un problème majeur, surtout dans certains pays où tout biais qui peut être démontré dans un outil d'évaluation peut entraîner des poursuites judiciaires. Toutefois, la preuve du biais, par opposition à la différence d'accomplissement, est souvent une question d'interprétation, liée au contexte social dans lequel l'évaluation a lieu.

Gipps et Murphy (1994) ont conclu leur livre intitulé *A Fair Test? Assessment, Achievement and Equity* en indiquant que le test impartial n'existe pas et ne peut exister, car la situation est trop compliquée et la notion trop simpliste. Cela ne veut toutefois pas dire que les concepteurs d'évaluations et les rédacteurs de sujets d'examen ne doivent pas faire tout ce qui est en leur pouvoir pour réduire l'incidence du biais et de l'injustice. Gipps et Murphy soutiennent également que les concepteurs d'évaluations doivent avoir pour but l'égalité des chances et l'égalité d'accès à l'évaluation plutôt que l'égalité des résultats, qui est fabriquée de toute pièce en manipulant chaque élément du test en fonction des statistiques obtenues pour sa réponse. Ils se demandent, par exemple, dans quelle mesure l'introduction de questionnaires à choix multiple dans les examens d'anglais afin d'améliorer les résultats des garçons serait justifiable, puisqu'elle fausserait la validité de l'évaluation selon la conception que l'on a de la définition de cette matière.

Il est largement reconnu que le manque d'impartialité dans le processus d'évaluation n'est qu'un des facteurs qui contribuent à l'iniquité dans le domaine de l'éducation, et peut-être même l'un des moins importants. Les différents résultats obtenus par les différents sous-groupes lors d'un test peuvent être le résultat de facteurs assez peu liés au test lui-même. En éducation, il existe bien d'autres sources d'iniquité qui ont un impact majeur sur les réalisations des candidats. On peut citer, par exemple, les différences de qualité de l'enseignement au sein d'un même établissement scolaire, les différences de ressources dans les différents établissements scolaires et les différentes régions géographiques, ainsi que les différences de situation sociale des candidats et les différences de niveau de soutien qu'ils reçoivent de la part de leur famille. Toutes ces sources, ou n'importe laquelle d'entre elles, pourraient compromettre considérablement les chances de réussite scolaire d'un candidat de telle façon que le processus d'évaluation, si impartial soit-il, ne pourrait rétablir l'équilibre. Par exemple, Smith et Tomlinson (1989) ont découvert que l'efficacité de l'enseignement dans un établissement scolaire était un facteur bien plus important pour expliquer les différences dans les résultats d'examen que l'origine ethnique des candidats. Cela montre que les tentatives qui sont faites pour adapter les outils d'évaluation afin d'empêcher que les résultats des différents groupes ethniques ne soient différents peuvent parfois être inappropriées.

Ce genre de considération a été utilisé pour justifier l'évaluation des aptitudes plutôt que celle de l'accomplissement, mais il est désormais reconnu qu'il est impossible d'évaluer l'aptitude, la capacité ou le potentiel seul, sans tenir compte du contexte social et du vécu scolaire. Il est également impossible d'examiner l'accomplissement scolaire de manière objective et indépendamment du contexte social et de la culture. Le concept de réussite scolaire est défini et mesuré en fonction des normes d'une partie restreinte de la société, quelle qu'elle soit.

Suppression du biais pour les candidats ayant des besoins en matière d'aménagement de la procédure d'évaluation

Un autre aspect du biais auquel il faut remédier est la possibilité qu'une tâche d'évaluation établisse une discrimination injuste contre les candidats ayant des besoins en matière de soutien à l'apprentissage dus, par exemple, à un diagnostic de dyslexie, de trouble déficitaire de l'attention ou de déficience visuelle. Il est possible d'y parvenir en veillant à ce que les conditions dans lesquelles les tâches d'évaluation sont effectuées tiennent compte de ces candidats, et ce, de manière appropriée, pour qu'ils puissent montrer leur accomplissement scolaire sur un pied d'égalité avec les autres candidats.

Ce thème est davantage développé dans la section « [L'équité pour tous : répondre aux besoins des candidats](#) ».

Reconnaissance du biais

Il ne faut pas perdre de vue que le biais peut aussi bien être positif que négatif. Si une tâche est très familière à un groupe de candidats ou qu'il leur est plus facile qu'à d'autres de la réaliser, il s'agit bien d'un biais. Une évaluation juste a pour objectif de donner la même chance à tous les candidats.

Il a été expliqué de quelle façon un biais peut être introduit dans la conception des évaluations ou dans le processus de notation. Bien qu'il soit très important d'être proactif en abordant le biais dans ces phases du cycle d'évaluation, une simple réflexion sur le biais potentiel ne suffit pas. Il est en effet nécessaire de chercher des preuves du biais dans les résultats du candidat, en comparant les résultats obtenus pour une question en particulier à ceux obtenus à l'examen dans son ensemble.

Il est également important de fonder les décisions relatives au biais sur des preuves, et non sur des stéréotypes. Cette remarque vaut tout particulièrement pour les différences entre les sexes, étant donné que certaines croyances très largement répandues au sujet des questions favorisant les garçons ou les filles peuvent se vérifier ou non pour la cohorte qui passe les examens de l'IB. Des preuves doivent systématiquement être tirées des évaluations précédentes avant d'émettre le moindre jugement à ce sujet.

Comparabilité

La comparabilité constitue l'un des aspects les plus complexes de la validité. Les résultats de l'évaluation sont régulièrement utilisés pour comparer les candidats à des fins de sélection. Lorsque deux candidats passent le même examen en même temps, il est possible d'affirmer avec certitude que le candidat ayant obtenu une note finale de 7 a mieux réussi ce jour-là que celui ayant obtenu une note finale de 4. Cependant, des comparaisons plus complexes sont souvent faites, comme dans les exemples qui suivent.

- Deux candidats ayant obtenu une note finale de 6 en histoire, mais ayant répondu à des questions différentes ou choisi des options différentes.
- Deux candidats, l'un ayant obtenu une note finale de 5 en littérature espagnole en mai 2014 et l'autre ayant obtenu une note finale de 5 dans cette matière en novembre 2012.
- Deux candidats, l'un ayant obtenu une note finale de 4 en physique et l'autre ayant obtenu une note finale de 4 en chimie.
- Deux candidats, l'un ayant obtenu une note finale de 4 en mathématiques et l'autre ayant obtenu une note finale de 4 en géographie.
- Un candidat européen ayant obtenu une note finale de 3 en informatique et une note finale de 6 en littérature chinoise, et un candidat africain ayant obtenu une note finale de 5 en littérature japonaise et une note finale de 4 en biologie.
- Deux candidats de 15 ans, l'un ayant obtenu le certificat du PEI et l'autre s'étant présenté aux examens d'un autre organisme décernant des titres.
- Deux candidats, l'un ayant obtenu une note finale de 6 en indonésien B au NM et l'autre ayant obtenu une note finale de 5 en indonésien B au NS.

La comparabilité s'interroge sur la possibilité de mettre sur le même plan les résultats obtenus à deux évaluations. Cela semble particulièrement difficile entre les matières, car nous évaluons des choses différentes puis nous nous demandons si elles sont de valeur équivalente.

Le concept de la validité appliquée dans un but précis plutôt que comme la caractéristique d'une évaluation est particulièrement pertinent ici. Les résultats d'un candidat en musique sont une meilleure indication de sa préparation à devenir musicien professionnel que la même note finale obtenue en arts visuels, mais ces deux résultats peuvent avoir la même valeur pour déterminer sa préparation à étudier l'histoire.

Le problème de la comparabilité est rendu encore plus complexe par le fait que chaque candidat a ses propres points forts et points faibles, ce qui explique qu'un candidat trouve plus facile de réussir dans certaines matières que dans d'autres. L'IB a donc un grand défi à relever puisqu'il existe des disparités considérables dans les cohortes passant chaque examen, et ce, d'autant plus lorsqu'un choix de cours est proposé.

L'IB s'efforce de préserver les trois principes suivants concernant la comparabilité.

1. La qualité du travail associée aux différentes notes finales dans une matière ou une discipline est comparable d'une année à l'autre.

2. Les notes finales ont la même signification dans toutes les matières ou disciplines, de sorte que les différents parcours possibles pour obtenir le diplôme correspondant au programme (diplôme de l'IB, certificat du PEI, etc.) soient comparables.
3. Même si l'IB s'efforce de se concentrer sur les compétences de haut niveau, ses évaluations sont généralement comparables aux examens similaires proposés par les pays ou par d'autres organismes décernant des titres.

Mesure de la comparabilité

Il existe de nombreuses façons de mesurer dans quelle proportion deux évaluations sont comparables, et ce thème a fait l'objet de nombreux articles de recherche. Dans leur revue de littérature sur la comparabilité entre les matières, Coe *et al.* (2008) ont réparti les méthodes de comparaison des difficultés en deux grandes catégories : les méthodes statistiques et les méthodes fondées sur une appréciation.

Les méthodes statistiques s'attachent à comparer les résultats obtenus par les candidats aux évaluations et à chercher des tendances. Elles reposent sur l'idée que si deux évaluations sont comparables alors, avec un échantillon aléatoire de candidats suffisamment grand, les résultats devraient en moyenne être les mêmes aux deux évaluations.

À l'inverse, les méthodes fondées sur une appréciation font appel à des spécialistes des deux matières pour examiner les évaluations et donner un avis mûrement réfléchi sur leurs difficultés respectives. Divers outils et techniques de recherche sont utilisés pour garantir qu'ils comparent ce qui est comparable.

Ces deux approches présentent de graves défauts conceptuels. Dans leur article, Coe *et al.* ont ainsi formulé six critiques générales concernant ces techniques.

Critiques des méthodes statistiques	Critiques des méthodes fondées sur une appréciation
<ul style="list-style-type: none"> • Mesure de facteurs autres que la difficulté, comme l'enseignement et la motivation. • Caractère multidimensionnel : il peut n'y avoir aucun point commun entre les matières. • Problème de représentativité : les statistiques reposent-elles sur un groupe de candidats intrinsèquement biaisé ? • Différences entre les sous-groupes : si différents sous-groupes de candidats sont exposés à des niveaux de difficulté différents, cela ne remet-il pas en question les conclusions de la comparabilité relative ? • Manque de cohérence entre les méthodes : dès lors, est-il possible de considérer qu'une méthode est « correcte » ? • Problèmes liés à l'obligation d'égalité : quels en seront les effets sur les candidats qui passent les évaluations ? 	<ul style="list-style-type: none"> • Étendue des critères : pour pouvoir être appliqués à différentes matières, les critères doivent être très généraux, ce qui les rend imprécis. • Notation des réponses en fonction du niveau de difficulté des questions : les examinateurs ont tendance à accorder davantage de points aux bonnes réponses apportées aux questions faciles qu'aux réponses plus faibles apportées aux questions plus difficiles. • Notation des différents types de travaux : les examinateurs ont beaucoup de mal à comparer des tâches différentes (par exemple, les réponses brèves et les dissertations). • Appui sur des comparaisons statistiques : ces méthodes sont influencées par les connaissances en matière de schéma d'évolution probable des résultats de candidats types. • Interprétation et contexte : une comparaison est établie entre un examen final unique et une série d'évaluations modulaires. • Regroupement d'appréciations : la plupart des évaluations mesurent plusieurs critères qui doivent être équilibrés.

Les partisans de l'une des méthodes sont souvent très critiques à l'égard de l'autre méthode, et d'autres soutiennent que toutes les approches actuelles sont fondamentalement erronées. Pour étudier ce sujet, Coe *et al.* ont appliqué cinq méthodes de mesure de la comparabilité entre les matières au GCSE (General Certificate of Secondary Education) et au GCE A Level (General Certificate of Education Advanced Level). Ils en ont conclu qu'il existait un niveau de cohérence relativement élevé entre les mesures de la comparabilité et que les différences qui les séparaient étaient bien moins importantes que les différences entre les matières. Ils ont également constaté que les difficultés dans chaque matière restaient stables d'une année à l'autre.

L'IB assure la comparabilité entre les années et les options grâce à une méthode passant par l'appréciation de l'examineur, l'analyse statistique et le retour d'information de l'enseignant (voir la section « [L'attribution des notes finales \(et la totalisation\)](#) »). La comparabilité est également passée en revue sur le plan des matières ou des disciplines à l'aide d'approches statistiques (paires de matières, simultanéité des accomplissements) et d'avis de spécialistes.

Approche de la validité adoptée par l'IB

- L'IB accorde plus d'importance à l'authenticité et à la pertinence de construit de l'évaluation qu'à la recherche de la fiabilité maximale.
- L'IB croit en une éducation globale et équilibrée. Sa priorité est de construire une argumentation solide qui montre la validité de chaque programme. Cette validité prévaut sur celle des cours et options pris individuellement.

La validité consiste en une recherche complexe et multidimensionnelle d'un équilibre entre diverses exigences importantes et contradictoires. Il n'existe pas de réponse correcte : le choix du point d'équilibre résulte d'un jugement fondé sur les valeurs de l'organisation élaborant les évaluations.

Pour l'IB, il est fondamental d'évaluer ce qui est important et de le faire de manière authentique en proposant des évaluations qui tiennent compte du monde réel. La notion de « pertinence de construit » englobe le premier point, car les évaluations de l'IB portent sur les éléments qui comptent vraiment dans chaque matière et non sur ce qu'il est facile de noter. Quant à l'authenticité, les évaluations de l'IB comportent des tâches riches de sens qui présentent des situations que les candidats peuvent rencontrer dans le monde réel, et non des tâches artificielles et forcées.

Ces objectifs sont cependant atteints au détriment de certains autres aspects de la validité, en particulier la fiabilité de l'évaluation. La notation des tâches authentiques et riches de sens s'accompagne généralement d'un degré important de subjectivité, ce qui implique d'accepter de plus grands écarts entre les notes des examinateurs qu'avec, par exemple, des questionnaires à choix multiple. Cela a également une incidence sur la capacité de gestion de l'évaluation. Ces types d'évaluations présentent en effet un plus grand défi et nécessitent plus de temps, que ce soit pour leur conception ou leur correction. Ils demandent également un plus grand investissement de la part des candidats en ce qui concerne la compréhension et le travail à fournir. Cela induit une charge de travail plus grande, facile à observer en comparant, par exemple, le travail fourni par un candidat lors de la réalisation de tâches nécessitant un travail de recherche sérieux au travail fourni pour se concentrer uniquement sur la réussite aux examens.

Bien que d'autres soient totalement libres de choisir des priorités différentes pour trouver un équilibre dans la validité de l'évaluation, l'IB est convaincu que sa position est adaptée et justifiable pour son système de vérification externe des évaluations.

L'IB a pour ambition de proposer des programmes d'études qui vont plus loin que les autres, en développant chez les jeunes la curiosité intellectuelle, les connaissances, la sensibilité et la détermination nécessaires pour réussir. Il espère que les élèves ainsi formés seront capables de bâtir un monde meilleur, dans un esprit d'entente mutuelle et de respect interculturel. Chacun de ses programmes contribue au développement des qualités du profil de l'apprenant de l'IB chez les élèves.

Du point de vue de l'évaluation, cela signifie que les buts de l'IB sont définis au niveau du programme et non au niveau des différentes matières ou disciplines. De ce fait, la question de la validité doit être posée au

niveau du programme et tenir compte des règles qui régissent l'attribution du certificat ou du diplôme, et pas simplement de chaque note finale.

Cela ne signifie pas pour autant qu'il n'est pas important de tenir compte de chaque cours ni de chaque tâche d'évaluation, et certains aspects de la validité n'ont en réalité de sens qu'à ce niveau. Cependant, construire un argument de validité au niveau global implique de réfléchir à l'ensemble du programme d'études suivi par l'élève.

La plupart des principes de l'évaluation exposés dans le présent document s'appliquent à l'ensemble du système éducatif de l'IB, mais la section C se penche sur chaque programme tour à tour et examine les particularités de chacun.

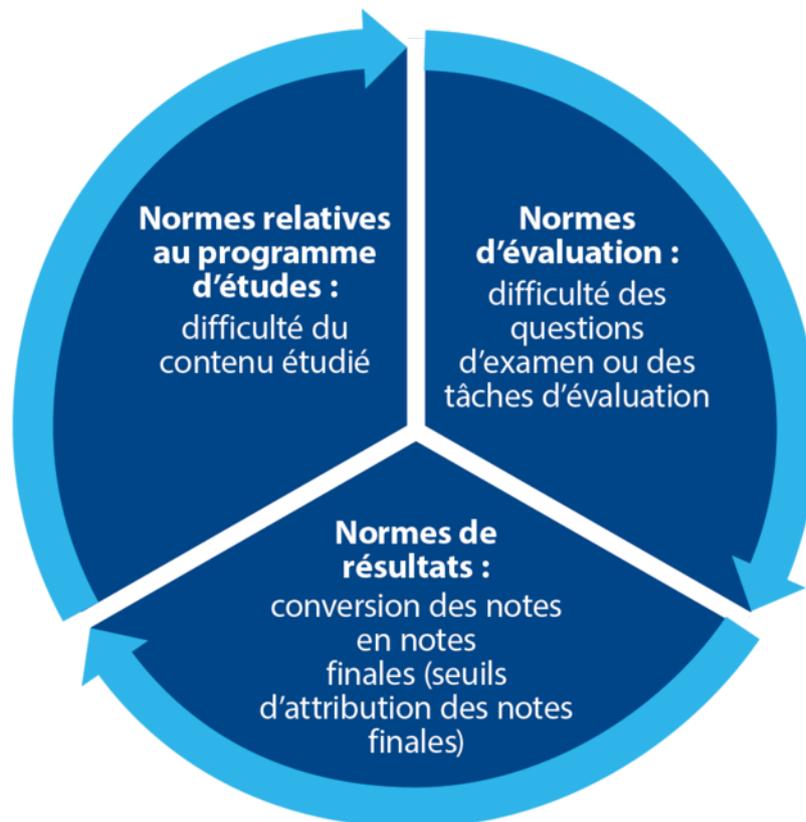
Définition des normes

- Toute définition des normes comporte trois volets : le programme d'études, l'évaluation et les résultats.
- Dans le cadre de l'évaluation formative, l'accent est habituellement mis sur les normes relatives au programme d'études et à l'évaluation.
- L'IB suit une « approche critériée limitée », c'est-à-dire qu'il utilise de manière équilibrée des critères et une comparaison avec les résultats obtenus par les candidats les années précédentes pour établir la norme.
- Il est tout aussi important de maintenir les normes que de les établir.

Les trois sens donnés aux normes

Dans le domaine de l'évaluation, les « normes » font généralement référence à la difficulté des tâches ou des tests conçus pour les candidats, et elles font partie intégrante de la comparabilité. Trois dimensions peuvent généralement être données aux normes.

Figure 18
Les trois aspects des normes



Si chaque aspect peut être modifié pour faire évoluer la difficulté générale de la matière, il ne faut pas perdre de vue que les modifications auront différents effets et seront assorties de différents délais.

L'IB modifie généralement un programme d'études lors du cycle de révision du programme, même si des modifications sont possibles à d'autres moments. À des fins d'équité, il convient de laisser aux enseignants le temps d'ajuster leur pédagogie pour qu'elle tienne compte des changements. Pour le Programme du diplôme et le POP, cela se fera dans l'idéal sur deux ans (la durée du programme), mais plus de temps pourra être nécessaire pour le PEI, étant donné qu'il dure cinq ans, même si l'évaluation ne porte en principe que sur les deux dernières années.

Les normes d'évaluation dépendent du cycle d'élaboration des épreuves, qui s'étale généralement sur 12 à 18 mois. Contrairement aux normes relatives au programme d'études, les normes d'évaluation varieront toujours légèrement d'une année à l'autre, car il est impossible de mettre au point deux séries de questions présentant le même niveau de difficulté. En effet, même si les questions sont reprises de l'épreuve précédente, les candidats seront en réalité exposés à un niveau de difficulté moindre, car ces questions utilisées pour la seconde fois leur seront, dans une certaine mesure, familières. Il peut arriver qu'une décision soit prise de changer délibérément les normes d'évaluation d'une épreuve si les épreuves précédentes n'ont pas produit les résultats escomptés et, en fonction du degré de changement nécessaire, il est possible que cela concerne une épreuve déjà en cours d'élaboration. Là encore, l'équité impose d'avertir les enseignants de ce changement, mais, selon son ampleur, il se peut que cet avertissement se fasse dans des délais bien plus courts que pour les changements relatifs au programme d'études.

Les normes de résultats (c'est-à-dire les seuils d'attribution des notes finales) sont ajustées chaque année pour suivre les modifications apportées aux normes d'évaluation. Par exemple, si l'épreuve est plus difficile et que les candidats obtiennent des notes plus basses, alors les seuils d'attribution des notes finales doivent être baissés en conséquence, afin que les candidats présentant un travail de même qualité obtiennent une note finale identique à celle qui leur aurait été attribuée l'année précédente. À l'inverse des deux cas précédents, ce sont ici surtout les examinateurs, et non les enseignants, qui doivent comprendre les normes de résultats, afin de pouvoir les maintenir chaque session. Une norme de résultats peut être modifiée lors de l'établissement des seuils d'attribution des notes finales, mais si l'IB a l'intention de le faire, il informe les établissements scolaires suffisamment à l'avance pour qu'ils puissent gérer et adapter en conséquence les attentes à l'égard des candidats. Dans un tel cas, il est primordial que les examinateurs comprennent cette nouvelle norme pour pouvoir l'appliquer lors des sessions ultérieures.

Au moment d'établir la norme globale pour une évaluation, il est important de trouver un équilibre entre ces trois définitions. Bien qu'il soit possible de poser des questions extrêmement difficiles sur un contenu simple ou de définir des normes de résultats extrêmement élevées pour une série de questions simples, il en résulte généralement une pertinence de construit très réduite. Par exemple, si le maximum de points est exigé à un test pour qu'une note finale de 7 puisse être attribuée, le candidat doit alors faire preuve de très hauts niveaux de précision. Un candidat ayant une excellente compréhension de la matière, mais étant peu soigneux ou montrant des faiblesses en matière de compétences de communication écrite aurait peu de chances d'obtenir une note finale de 7. S'agit-il là du but visé par l'évaluation ?

La définition de la norme s'applique à la fois à l'évaluation formative et à l'évaluation sommative. Dans le cas de l'évaluation formative, cependant, l'établissement des normes de résultats est habituellement intégré dans le jugement de l'enseignant, dans le retour d'information qu'il choisit de faire. En règle générale, on accorde bien plus d'attention aux normes relatives au programme d'études et aux normes d'évaluation afin de garantir que le test est adapté à l'apprenant et qu'il apportera des renseignements utiles pour orienter l'enseignement.

Figure 19

Il est peu probable que cette évaluation formative fournisse un retour d'information utile.



L'approche normative et l'approche critériée des normes de résultats

- Avec l'approche normative, la norme de résultats est établie à partir de la réussite des candidats. Il s'agit, par exemple, de tenir compte des résultats des 20 % de candidats les meilleurs.
- Avec l'approche critériée, la norme de résultats est établie à partir d'une description de ce qui devrait apparaître dans le travail accompli par le candidat.
- L'IB suit une « approche critériée limitée », c'est-à-dire qu'il utilise de manière équilibrée des critères et une comparaison avec les résultats obtenus par les candidats les années précédentes pour établir la norme.

Les termes « approche normative » et « approche critériée » désignent deux façons d'établir et de maintenir les normes de résultats des évaluations.

En quoi consiste l'approche normative ?

La définition technique de l'approche normative est généralement associée aux tests standardisés. Il s'agit d'essayer le test en le faisant passer à un échantillon type de candidats et de se servir des résultats (qui, par définition, devraient représenter une courbe de distribution normale ou en forme de cloche) comme d'une échelle de référence pour produire une note pour tout futur candidat qui passera ce même test. Le processus qui consiste à produire une distribution standard des notes à partir de l'essai initial constitue « l'étalonnage ».

Cette définition technique de l'approche normative n'implique pas nécessairement d'appliquer une distribution fixe à chaque ensemble de résultats au test ; la distribution fixe n'est utilisée que pour l'étalonnage initial. La distribution des notes qui seront obtenues par les futurs candidats peut être différente de cette distribution normale.

Dans la pratique, le terme « approche normative » désigne souvent le processus consistant à attribuer un rang aux candidats en fonction des résultats obtenus et à établir le pourcentage de candidats recevant chaque note finale. Par exemple, la meilleure note finale sera attribuée aux 15 % de candidats les meilleurs.

En quoi consiste l'approche critériée ?

Glaser (1963) a été le premier à proposer la notion d'évaluation critériée. Elle représentait un changement important dans la manière d'établir les normes de résultats, en mettant l'accent sur l'évaluation des accomplissements du candidat dans un domaine comportemental bien défini (Popham, 1978).

Dans l'approche critériée, les résultats du candidat sont comparés à une description prédéfinie des attentes correspondant à chaque note finale. Des spécialistes de la matière ou de l'évaluation se chargent en principe de cette comparaison, en faisant appel à leur jugement professionnel.

Cette approche présente des difficultés et des limites, car il est très difficile de proposer des descriptions sans équivoque qui ont la même signification pour tous les spécialistes amenés à émettre les jugements. En effet, il a été avancé qu'aucun critère, quel que soit le niveau de précision de sa formulation, ne permettait une interprétation sans équivoque (Wiliam, 1993). Le résultat d'un test critérié traditionnel est la démonstration ou non de la maîtrise du domaine pertinent.

Dans la pratique, chacune de ces deux approches comporte des inconvénients majeurs. Une approche normative stricte impose de prouver avec force que le test actuel présente le même niveau de difficulté que le test initial, alors que l'approche critériée est soumise à l'effet de Good et Cresswell (1988), selon lequel le jugement du spécialiste ne tient pas compte avec précision des exigences des questions.

Quelle est l'approche adoptée par l'IB ?

L'IB suit une approche critériée limitée, qui repose sur des critères et reconnaît aussi l'effet de Good et Cresswell. Cette approche demande aux examinateurs spécialisés d'établir un intervalle réduit dans lequel situer les seuils d'attribution des notes finales en partant des critères (descripteurs des notes finales). Une comparaison est ensuite faite avec les seuils calculés pour correspondre aux résultats des années précédentes. Lorsque ces deux seuils coïncident, les notes finales sont fixées, mais en cas de désaccord, d'autres discussions sont nécessaires pour déterminer comment harmoniser les preuves contradictoires.

Enfin, il convient de retenir que les tests critériés et normatifs diffèrent plus sur le plan de l'analyse et de l'interprétation des réponses des candidats que sur le plan des questions elles-mêmes.

Le maintien des normes

- Une fois les normes appropriées établies, l'IB doit s'assurer qu'elles sont appliquées chaque année.
- Les normes relatives au programme d'études sont revues lors de la révision du programme.
- Les normes d'évaluation et les normes de résultats sont maintenues en associant jugement professionnel et preuves statistiques.

La comparabilité est un aspect fondamental de la validité. Cela signifie qu'il faut veiller à ce que les mêmes normes soient appliquées chaque année. Il est important de comprendre que les normes de l'IB reposent sur le sens des notes finales, et non des notes. La différence entre ces deux types de notes est expliquée [ici](#).

Les normes relatives au programme d'études sont les plus faciles à maintenir, étant donné qu'elles restent les mêmes d'une session à l'autre. Les évaluations doivent être un reflet fidèle de l'ensemble du programme d'études, ce qui est vérifié lors de la rédaction des épreuves. Des changements peuvent toutefois être apportés aux normes relatives au programme d'études en raison de facteurs extérieurs. Les compétences en informatique illustrent parfaitement cette idée. Les ordinateurs évoluent et prennent une place grandissante dans la vie quotidienne. De ce fait, les connaissances et la compréhension qui étaient autrefois réservées aux spécialistes en raison de leur complexité se sont banalisées. Des modifications ont donc dû être apportées à la norme relative au programme d'études portant sur ce thème, même si le contenu reste inchangé.

Pour résoudre les problèmes de ce type, mais aussi pour garder le contenu à jour, l'IB s'appuie sur un cycle de révision des programmes. Il est prévu que les normes relatives aux programmes d'études soient modifiées à l'issue de ces révisions, et un équilibre doit alors être trouvé en ce qui concerne les normes d'évaluation et les résultats afin de ne pas désavantager les candidats.

Concernant les examens de l'IB, les normes d'évaluation sont modifiées chaque session, car il est impossible que deux épreuves d'examen présentent exactement les mêmes exigences. Dans certains systèmes éducatifs, toute une série de tests préliminaires sont organisés pour déterminer la difficulté de chaque question, et les épreuves d'examen sont soigneusement conçues afin de garantir un niveau de difficulté adapté. L'IB a fait le choix de ne pas procéder à ce type de tests préliminaires, en raison du risque de divulgation des questions avant que les candidats ne passent les examens. La conception d'épreuves cohérentes et équilibrées repose donc sur le jugement professionnel et l'expérience des auteurs d'épreuves et des vérificateurs. Les normes de résultats sont ensuite ajustées à partir des réponses des candidats pour veiller au maintien de la norme globale.

Les deux principales méthodes de maintien des normes de résultats sont examinées dans la section précédente, et le processus d'attribution des notes finales a pour but de maintenir ces normes.

Description de la réussite : les résultats des candidats à l'évaluation sommative

- L'objectif affiché par la déclaration de mission de l'IB est de former des jeunes capables de bâtir un monde meilleur, et une évaluation réussie devra soutenir cet objectif.
- Même si les notes finales ne sont qu'un aperçu très simplifié des accomplissements des candidats, elles permettent aux interlocuteurs tels que les universités, les employeurs ou les divers établissements scolaires de disposer d'une base pour décider de leur sélection.
- Si l'IB ne fournissait que des informations plus complexes et globales, d'autres auraient à simplifier ces preuves pour prendre des décisions éclairées en vue d'une sélection, ce qui serait certainement fait avec une attention moins aiguë que celle apportée par l'IB lors de l'établissement des seuils d'attribution des notes finales.
- Dans ses évaluations, l'IB considère que le jugement professionnel est important pour atteindre un résultat riche de sens pour le candidat, mais il reconnaît également que ce jugement doit être étayé par des preuves objectives afin de minimiser le biais.
- Les résultats doivent être suffisamment précis pour ne pas désavantager les candidats en dressant un bilan qui ne serait pas un bon reflet de leur accomplissement. La priorité reste cependant de concevoir des évaluations utiles.

La tyrannie des notes finales : un moindre mal

Réfléchissez aux connaissances, aux compétences et à l'expérience qui font l'excellence d'un chef. Ses connaissances portent sur les combinaisons d'ingrédients et de saveurs qui lui permettront de composer un plat parfaitement équilibré. Ses compétences interviennent pour le choix et la préparation des ingrédients, la détermination du mode de cuisson (plaque chauffante ou four) et la présentation. Et son expérience concerne l'utilisation des techniques pour atteindre la perfection, l'appréciation d'un plat merveilleusement cuisiné et la présentation d'un ensemble d'ingrédients et de plats pour procurer une immense satisfaction à table.

Figure 20

Connaissances, compétences et expérience



Réduisez maintenant toute cette complexité en une note finale unique sur 7 pour déterminer qui est le meilleur chef. Le résultat est presque vide de sens. Serait-il utile d'augmenter l'échelle et de choisir une note finale sur 100, voire sur 1 000 ? La réponse sera probablement « non ». Cela permettrait un plus grand champ de différenciation, mais le problème fondamental, à savoir comparer des ensembles de compétences différents, ne serait pas résolu.

Ce problème est précisément celui que pose l'évaluation : comment représenter la complexité des connaissances, de la compréhension et des capacités des apprenants en un résultat unique qui en ferait la synthèse ? Même si toutes les informations sur un candidat pouvaient être recueillies, cela ne signifierait pas pour autant qu'il serait possible de lui attribuer une note finale reflétant parfaitement ses talents.

Une autre possibilité, souvent suggérée, serait de ne pas attribuer de note finale du tout, mais peut-être de fournir pour chaque candidat une description personnalisée de ce qui a été bien réussi et des aspects à renforcer. Une telle approche respecte les principes d'un enseignement et d'un apprentissage de qualité, mais elle présente l'inconvénient majeur (que certains de ses partisans considèrent néanmoins comme une force) de compliquer considérablement les comparaisons entre candidats.

Pour tenter de résoudre ce problème, il est utile de revenir au but des évaluations de l'IB, qui est d'aider les candidats à pouvoir poursuivre leurs études ou à intégrer le monde du travail. Cela signifie qu'une forme de sélection devra être pratiquée par les établissements d'accueil (généralement les universités pour le Programme du diplôme ou le POP et les établissements scolaires pour le PEI) afin de choisir les candidats qui seront acceptés.

À l'évidence, si les autres critères sont moins fiables que les examens, compter davantage sur ces critères signifiera faire reposer la sélection sur des décisions moins fiables.

(Cresswell, 1986, p. 37 – 54)

Cette citation figure dans l'introduction du présent document, mais elle mérite d'être rappelée. Si une sélection doit avoir lieu, alors l'IB est tenu d'apporter son soutien aux élèves en rendant ces décisions en matière de sélection aussi justes et pertinentes que possible. Si les établissements d'accueil ne se voient remettre que des bilans descriptifs pour chaque élève, ils auront à trouver une autre méthode de comparaison, qui sera certainement moins fiable et uniforme que les résultats aux examens exprimés sous forme de notes finales. Réfléchissez, par exemple, à la validité d'un court entretien avec un tuteur et à tous les facteurs susceptibles d'influencer le résultat qui ne devraient pas servir de base pour la sélection.

L'intention n'est pas de laisser entendre que les évaluations sommatives sont un moyen idéal ou particulièrement bon pour procéder à une telle sélection, mais elles sont plus justes que les autres possibilités. Plus important encore, l'IB s'efforce constamment de rendre cette méthode la plus juste, utile et fiable possible.

Figure 21

Une distinction devrait-elle être faite entre les deux candidats de chaque groupe ?



Dans l'exemple ci-dessus, la différence entre les deux candidats du premier groupe est celle qui est faite entre les notes finales de 5 et 6 dans les descripteurs généraux des notes finales du Programme du diplôme. Si vous deviez choisir entre ces deux candidats, votre décision serait probablement juste, et ces candidats obtiendraient sans doute le même résultat s'ils devaient repasser le test. Dans le second groupe, la différence risque de ne pas être significative, et les deux candidats ont probablement la même aptitude.

Cet exemple prouve que toutes les différences ne sont pas significatives et que l'utilisation des notes finales donne une indication des aspects qui permettent de distinguer deux candidats. Il est certainement vrai que pour un candidat se situant à la limite entre deux notes finales (que ce soit juste au-dessus ou juste au-dessous du seuil d'attribution) n'importe laquelle des deux notes finales est juste, mais, pour la plupart des candidats, une différence entre les notes finales représente un écart d'accomplissement important.

Ce raisonnement amène à un débat sur le nombre de notes finales à employer. Avec seulement deux notes finales (réussite ou échec), la plupart des candidats auraient une note juste, mais les conséquences seraient sérieuses pour ceux qui se situent à la limite. À l'inverse, avec 20 notes finales, bien plus de candidats se situeraient à une limite, mais, pour chacun, les conséquences d'un rattachement à la mauvaise catégorie seraient moins graves. Ce concept a fait l'objet d'une étude plus poussée par Cresswell (1986).

En général, l'IB a retenu sept notes finales pour représenter le nombre de catégories importantes découlant de ses évaluations et pour trouver un juste équilibre entre le nombre de candidats se situant à une limite et les conséquences de l'attribution d'une note finale inappropriée.

L'importance du jugement professionnel

Les compétences de pensée complexe et de haut niveau sur lesquelles se concentrent les évaluations de l'IB ne se prêtent pas facilement à une notation simple et exempte de jugement. Les réponses des candidats seront probablement très variées, tout en étant également valables et correctes. La recherche semble indiquer que les connaissances et les compétences complexes ne devraient pas être décomposées en petits éléments distincts pour être enseignées, et ce principe vaut également pour la notation de ces réponses. Lorsque les barèmes de notation sont élaborés, des orientations très précises doivent être données quant à la manière de noter le travail de tous les candidats, et ce, de façon homogène.

Cela implique d'accorder une grande importance au jugement professionnel des correcteurs, et surtout à l'expérience professionnelle des examinateurs superviseurs qui établissent et expliquent les normes de notation. Il s'agit là d'un grand défi pour la fiabilité du système d'évaluation, mais ce défi doit être relevé dans le contexte de la précision des résultats.

Notation des évaluations

- Le terme « notation » désigne le processus consistant à évaluer dans quelle mesure un candidat a bien réalisé la tâche confiée.
- Même s'il est courant de percevoir la notation comme le fait de déterminer si une réponse est « correcte » et d'attribuer une note en conséquence, il ne s'agit que d'une des nombreuses visions possibles. La notation peut se concentrer sur chacun des détails d'une réponse ou au contraire porter un jugement plus global.
- D'autres approches de la notation existent, notamment le jugement comparatif.
- Dans le cadre de l'évaluation formative, la notation peut ne pas produire de résultat numérique et être uniquement descriptive.

Qu'entend-on par « notation » ?

Figure 22

Raisons pour lesquelles les notes et les résultats sont deux choses différentes



La notation ne décrit pas dans quelle mesure le candidat a réussi. Le niveau atteint dépend de plusieurs facteurs, dont le degré de difficulté des questions et la « note de passage ». La notation consiste à comparer

la réponse du candidat à la réponse parfaite. Dans la figure ci-dessous, l'idée selon laquelle un chameau est un cheval décrit par un comité illustre bien le fait que des critères (ou un barème de notation) inappropriés peuvent produire un mauvais résultat.

Figure 23

Il peut être nécessaire de revoir la réponse parfaite après examen des travaux des candidats.

Sujet d'examen : Consigne – Représentez un cheval.



Examineur : « Je crois qu'on a besoin d'un nouveau barème de notation ! »

Il est possible de procéder à la notation de multiples façons, en fonction de la nature de la tâche. Il arrive que la notation soit très objective – la réponse du candidat est soit correcte, soit incorrecte –, et c'est souvent le cas lorsque seuls quelques mots-clés sont à donner ou lorsque le candidat doit faire un choix parmi plusieurs réponses possibles.

Mais elle peut aussi être bien plus subjective, lorsqu'elle demande au correcteur de juger si le candidat a produit une réponse acceptable ou de choisir, parmi plusieurs énoncés (appelés « bandes de notation »), celui qui décrit le mieux la mesure dans laquelle la réponse du candidat correspond à la réponse parfaite. Des exemples de bandes de notation utilisées dans le cours de gestion des entreprises du Programme du diplôme et le cours d'éducation physique et à la santé du PEI sont fournis ci-après.

Gestion des entreprises – Critère B : mise en application

Ce critère permet de déterminer dans quelle mesure l'élève est capable d'appliquer les outils, techniques et théories pertinents de gestion des entreprises à l'organisation faisant l'objet de l'étude de cas.

Points	Descripteurs de niveaux
0	Le travail n'atteint pas l'un des niveaux décrits ci-dessous.
1	Les outils, techniques et théories pertinents de gestion des entreprises sont reliés à l'organisation faisant l'objet de l'étude de cas, mais ce lien est inapproprié ou superficiel.
2	Les outils, techniques et théories pertinents de gestion des entreprises sont reliés de façon appropriée à l'organisation faisant l'objet de l'étude de cas, mais ce lien n'est pas développé.
3	Les outils, techniques et théories pertinents de gestion des entreprises sont généralement bien appliqués pour expliquer la situation et les problèmes de l'organisation faisant l'objet de l'étude de cas. Toutefois, cette explication peut manquer d'approfondissement ou d'envergure. Des exemples sont fournis.

Points	Descripteurs de niveaux
4	Les outils, techniques et théories pertinents de gestion des entreprises sont bien appliqués pour expliquer la situation et les problèmes de l'organisation faisant l'objet de l'étude de cas. Les exemples sont appropriés et explicatifs.

Éducation physique et à la santé (5^e année du Programme d'éducation intermédiaire) – Critère A : connaissances et compréhension

Figure 24

Exemples de notations subjectives dans le Programme du diplôme et le PEI

Niveau	Descripteurs de niveaux
0	L'élève n'atteint aucun des niveaux décrits ci-dessous.
1 – 2	L'élève : <ul style="list-style-type: none"> • indique des connaissances factuelles, procédurales et conceptuelles relatives à l'éducation physique et à la santé ; • applique des connaissances relatives à l'éducation physique et à la santé pour rechercher des questions et suggérer des solutions à des problèmes tirés de situations familières ; • applique la terminologie spécifique à l'activité physique et à la santé afin de transmettre sa compréhension, mais avec un succès limité.
3 – 4	L'élève : <ul style="list-style-type: none"> • résume des connaissances factuelles, procédurales et conceptuelles relatives à l'éducation physique et à la santé ; • applique des connaissances relatives à l'éducation physique et à la santé pour analyser des questions et résoudre des problèmes tirés de situations familières ; • applique la terminologie spécifique à l'activité physique et à la santé afin de transmettre sa compréhension.
5 – 6	L'élève : <ul style="list-style-type: none"> • identifie des connaissances factuelles, procédurales et conceptuelles relatives à l'éducation physique et à la santé ; • applique des connaissances relatives à l'éducation physique et à la santé pour analyser des questions et résoudre des problèmes tirés de situations aussi bien familières que nouvelles ; • applique la terminologie spécifique à l'activité physique et à la santé de manière cohérente afin de transmettre sa compréhension.
7 – 8	L'élève : <ul style="list-style-type: none"> • explique des connaissances factuelles, procédurales et conceptuelles relatives à l'éducation physique et à la santé ; • applique des connaissances relatives à l'éducation physique et à la santé pour analyser des questions complexes et résoudre des problèmes complexes tirés de situations aussi bien familières que nouvelles ; • applique la terminologie spécifique à l'activité physique et à la santé de manière cohérente et efficace afin de transmettre sa compréhension.

Une autre variable concerne la façon dont la notation est réalisée pour plusieurs aspects différents du travail, qui portent souvent le nom de « critères ». Par exemple, une dissertation pourrait être évaluée selon quatre critères distincts : (1) la qualité de la grammaire, (2) l'exactitude des faits essentiels, (3) la structure de la dissertation et (4) la qualité de la conclusion. Pour que cette approche soit réussie, il faut s'assurer que les critères sont indépendants les uns des autres. Dans l'exemple donné précédemment, si un candidat ne faisait pas de conclusion, il pourrait être pénalisé dans les critères 3 et 4.

Le contraire de la notation à l'aide de critères est le jugement reposant sur l'impression générale. Dans ce cas, le correcteur trouve un compromis entre les différents aspects d'une réponse idéale et émet un jugement final qui reflète le travail dans son ensemble.

La notation globale et la notation à l'aide de critères

Les deux approches présentent divers avantages et inconvénients. Avec l'impression générale, il est souvent difficile pour deux examinateurs différents de parvenir à une note cohérente, car ils doivent trouver un compromis entre les différents aspects du travail. Cependant, le dernier classement (c'est-à-dire la comparaison des différents travaux de candidats) reflète souvent fidèlement le meilleur travail.

À l'inverse, l'utilisation de critères permet une notation bien plus cohérente, même si de petites différences peuvent facilement prendre de l'ampleur. Par exemple, s'il y avait trois critères, chacun noté sur 4, et qu'un candidat se situait entre 2 et 3 pour chacun, alors deux examinateurs sensés pourraient lui attribuer la note 6 (2 à chaque critère), la note 9 (3 à chaque critère) ou toute note intermédiaire, ce qui pose un problème relativement grave étant donné que la note maximale est de 12 seulement.

L'autre difficulté présentée par la notation à l'aide de critères est qu'elle peut parfois produire des résultats corrects qui ne semblent pas justes. C'est notamment le cas lorsqu'un candidat a très bien réussi à faire correspondre son travail aux critères et qu'un autre candidat a effectué un « bon » travail sans le faire bien correspondre aux critères.

Les autres formes de notation

Pour beaucoup, la notation d'un travail ressemble à l'image qui suit.

Figure 25

Perception typique de la notation



Il ne s'agit pourtant pas de la seule façon d'aborder la notation, et une approche a fait l'objet de nombreuses études récentes : le jugement comparatif.

Le jugement comparatif repose sur l'idée que l'esprit humain est mieux à même de comparer deux objets que d'émettre des jugements à partir d'une échelle abstraite. À titre d'exemple, observez la première image

qui suit. Sur une échelle d'évaluation de la chaleur allant de 1 à 10 (10 illustrant une chaleur extrême), quelle note lui attribueriez-vous ? Comparez maintenant la première image à la seconde. Laquelle des deux images représente l'élément le plus chaud ?

Figure 26

Qu'est-ce qui est le plus chaud ? Le café ou le volcan ?



Même si cet exemple est simpliste, la démonstration qu'il met en avant reste vraie pour toutes les décisions subjectives, comme déterminer dans quelle mesure un candidat a bien répondu à la question de l'examen dans sa dissertation plutôt que comparer deux dissertations pour trouver celle qui a le mieux traité la question.

L'approche qui sous-tend le jugement comparatif est que les examinateurs prennent beaucoup de décisions pour déterminer « le mieux et le moins bien » (une opposition souvent exprimée par les termes « gagnant » et « perdant ») et que ces décisions sont ensuite combinées pour créer un ordre de classement des éléments évalués. En s'appuyant sur les mathématiques, il est possible non seulement de gérer des cas imprévus (A gagne contre B, B gagne contre C, mais contre toute attente, un juge décide que C gagne contre A), mais encore d'établir l'ampleur de l'écart qui doit séparer deux éléments dans un classement en estimant la vraisemblance de chaque résultat (gagnant ou perdant). Ce raisonnement peut alors être utilisé pour trouver les notes traditionnelles correspondantes.

Lorsqu'il y a plusieurs juges, ce qui importe le plus dans le jugement comparatif est que tous doivent se mettre d'accord sur la définition d'une « bonne » réponse. Cela peut apparaître de façon évidente dans de nombreuses situations, comme avec l'exemple de la chaleur ci-dessus, mais lorsque cela détermine la note du travail des candidats, il est important d'être explicite. En histoire, par exemple, une dissertation est « bonne » si elle comporte une argumentation convaincante, étayée par des informations précises, plutôt que si elle est très longue et pleine de faits sans rapport les uns avec les autres. Cette description d'un « bon » travail est appelée « déclaration d'importance ».

Le second aspect du jugement comparatif est que les « notes » définitives reposent sur les jugements de tous les examinateurs ayant corrigé la dissertation dans le cadre du processus, et pas sur le jugement d'un seul examinateur. De ce fait, les notes sont issues d'un consensus trouvé entre les visions de tous les examinateurs, et non d'un examinateur unique. Cette approche est très différente de celle qui est habituellement suivie pour l'évaluation et exige une nouvelle interprétation du sens donné à la « fiabilité ».

Le plus gros inconvénient de cette méthode est qu'elle nécessite bien plus de décisions en matière de notation. Au lieu d'être examiné une seule fois, chaque élément demande plusieurs analyses, même si elles se font généralement plus rapidement. Une notation réalisée à l'aide du jugement comparatif prend donc du temps et, pour régler ce problème, une variante connue sous le nom de « jugement comparatif adaptatif » s'est développée. Elle vise à obtenir un résultat cohérent à partir d'un nombre moins important de décisions « gagnant/perdant », en recherchant les jugements qui sont vraiment nécessaires.

Le jugement comparatif est le plus susceptible d'être avantageux pour l'IB quand des tâches authentiques et importantes sont difficiles à noter de façon fiable. Pour de plus amples informations sur le jugement comparatif et sur le jugement comparatif adaptatif, veuillez consulter la section « Bibliographie ».

La notation et l'évaluation formative

La notation n'a pas à être uniquement numérique. Il est tout à fait possible de comparer un travail à la réponse « parfaite » et de fournir des commentaires descriptifs sur les similitudes et les différences entre les deux. Cela rend toutefois très difficile toute comparaison entre deux réponses afin de déterminer celle qui est la meilleure, mais avec l'évaluation formative, qui s'attache à fournir un retour d'information pour soutenir l'apprentissage, une telle comparaison n'est pas nécessaire.

Qu'est-ce qu'une bonne évaluation ?

- Il n'existe pas de réponse unique à cette question. Tout dépend de l'importance accordée aux différentes priorités et du but de l'évaluation.
- La démarche de l'IB repose sur le principe qu'il faut évaluer ce qui est important et non considérer comme important ce qui peut être évalué.
- Les évaluations de l'IB cherchent à avoir un effet de remous positif sur l'enseignement et sur l'apprentissage.
- En règle générale, les évaluations doivent inclure un éventail de tâches et permettre d'organiser des activités plus poussées en classe ainsi que des examens.
- Malgré les nombreux paramètres techniques à prendre en considération pour définir le sens de la validité d'une évaluation, cette définition doit avoir pour valeur centrale qu'il s'agit d'une bonne évaluation des buts visés par l'évaluation.

Il est en réalité très difficile de répondre à cette question toute simple. Cela s'explique par le fait que les priorités sont différentes en fonction de chacun et, donc, qu'il n'y a pas de bonnes ni de mauvaises réponses.

La démarche de l'IB repose sur le principe qu'il faut **évaluer ce qui est important** et non considérer comme important ce qui peut être évalué. Ces éléments doivent être pondérés avec tous les autres aspects, comme la fiabilité et la charge de travail des candidats.

Il est important de comprendre qu'il est difficile pour chaque approche de satisfaire à toutes les priorités possibles. Plus précisément, une bonne conception de l'évaluation est différente pour l'évaluation sommative et l'évaluation formative. En partant de ce principe, il est possible de résumer comme suit les aspects que l'IB associe à une bonne évaluation :

- soutien des objectifs pédagogiques ;
- utilisation d'un éventail de tâches d'évaluation ;
- prise en considération des compétences plus larges du candidat et des compétences de pensée de haut niveau.

Chacun de ces aspects sera abordé à tour de rôle.

Une bonne évaluation soutient les objectifs pédagogiques

- L'évaluation doit encourager les bonnes pratiques d'enseignement (remous positif).
- L'évaluation doit être prévisible.

L'évaluation ne doit pas être perçue comme une activité indépendante de l'enseignement et de l'apprentissage. Les résultats de l'évaluation de l'IB reposent sur une évaluation sommative, et ils n'ont pas vocation à fournir un retour d'information direct sur l'enseignement et l'apprentissage. Il est toutefois bien entendu que ce qui est inclus dans l'évaluation aura un impact sur ce qui est enseigné. Il s'agit de l'effet de remous.

L'IB part du principe que la conception des évaluations doit encourager les résultats pédagogiques les plus souhaitables pour les élèves. Les effets sur l'apprentissage des élèves demeurent un aspect fondamental de la conception des évaluations et constituent, avec la pertinence de construit (c'est-à-dire le fait d'évaluer les

bonnes choses), l'une des priorités de l'IB pour déterminer comment pondérer les différents éléments de la validité.

L'évaluation a des enjeux importants et exerce une influence considérable sur l'enseignement et sur l'apprentissage. Il est possible d'en tirer parti en concevant des outils d'évaluation qui encouragent une bonne méthode d'enseignement et une implication constructive des élèves dans leur propre apprentissage, tout en tenant compte des opinions récentes en matière de théorie de l'apprentissage (par exemple, Murphy, 1999).

Les caractéristiques personnelles que l'on cherche à développer chez les élèves, et qui sont exprimées dans la déclaration de mission de l'IB, s'accordent très bien avec la théorie constructiviste de l'apprentissage des élèves dans laquelle ces derniers s'engagent activement dans le processus d'apprentissage, assument la responsabilité de leur propre apprentissage et approfondissent leurs connaissances, leur compréhension et leurs compétences par l'intermédiaire de la recherche.

Dans un certain nombre de matières, les modalités d'évaluation requièrent des élèves qu'ils montrent une appréciation de perspectives culturelles différentes des leurs. L'altruisme et la compassion étant des qualités plus affectives, il est plus difficile d'en tenir compte dans l'évaluation officielle ; elles doivent néanmoins être représentées dans le système d'évaluation global. Cela est en partie rendu possible grâce aux éléments du programme d'études qui ne sont pas évalués, comme le service communautaire dans le PEI ou le programme créativité, activité, service (CAS) dans le Programme du diplôme et le POP. Cela étant, les évaluations de l'IB illustrent aussi des pratiques de travail éthiques ainsi qu'une compréhension et une valorisation des différences.

Pour ce qui est de la conception des cours, l'IB accorde une grande importance à la validité prédictive (la mesure dans laquelle les résultats peuvent prédire la réussite ultérieure), en sachant que la manière dont l'évaluation est menée influence grandement la façon dont les cours de l'IB sont enseignés dans les établissements scolaires. Le modèle d'évaluation (ensemble d'outils d'évaluation) utilisé dans chaque matière est conçu de manière à avoir une large base, et notamment divers types de preuves. Le but est en effet de favoriser la pertinence de construit en fournissant un éventail de preuves aussi large que possible pour soutenir les résultats et l'apprentissage des élèves.

Alors que l'IB est parfaitement conscient de l'effet de remous exercé par les évaluations sur l'enseignement et l'apprentissage, il encourage les établissements scolaires à adopter une pédagogie développant l'intégralité des buts et des principes philosophiques de ses programmes chez les élèves.

L'IB mène régulièrement des études pour évaluer dans quelle mesure les programmes conçus constituent une bonne préparation à la poursuite de la scolarité. Pour de plus amples informations sur les études réalisées par l'IB, veuillez vous reporter à la section dédiée à la [recherche](#) sur son site Web.

Qu'est-ce qu'une bonne prévisibilité ?

Le terme « prévisibilité » désigne ici la capacité à évaluer ce qui se passera ou à quel moment une chose se passera. Dans le domaine de l'évaluation, cela fait référence à l'aptitude des établissements scolaires à déterminer les questions qui seront posées dans une épreuve et le moment auquel elles seront posées. Il est fondamental que les pratiques de travail de l'IB en matière d'évaluation comportent une bonne prévisibilité, car, en adhérant à ce principe, l'IB indique qu'il respecte les exigences de ses construits, qui sont publiées à l'intention des enseignants, ce qui conduit à une possibilité d'évaluation « juste » du point de vue de l'alignement avec le programme d'études. Ce que l'IB a annoncé vouloir évaluer sera évalué.

Pour proposer des options d'enseignement, les établissements scolaires doivent fournir un investissement, et l'IB veut s'assurer qu'ils en sont récompensés pendant tout le cycle de vie d'un programme d'études (avant qu'il soit révisé). Tout le programme doit être examiné comme l'exige l'évaluation. Une grande prudence est montrée pour supprimer les risques de mauvaise prévisibilité à la fin d'un cours (lorsqu'un établissement se rend compte de ce qui n'a pas encore été demandé et qui apparaîtra donc probablement dans une épreuve).

Le principe sous-jacent est que rien ne doit constituer une surprise, ni pour le candidat ni pour l'établissement scolaire. Quelle que soit la composante concernée, les questions posées doivent être explicitement inspirées du guide pédagogique correspondant. Dans la mesure du possible et s'il y a lieu, l'IB vise à éliminer les problèmes dus à la prévisibilité en évaluant les compétences au moyen d'évaluations

conçues spécialement pour ne pas avantager les candidats disposant de réponses préparées (par exemple, il cherche un équilibre entre les connaissances, la compréhension et l'étude d'un stimulus jamais vu auparavant).

La conception de l'évaluation s'avère primordiale pour veiller à ce que les façons d'évaluer tout type de thème, d'option ou de texte soient suffisantes, afin de limiter la mauvaise prévisibilité.

Lorsque le modèle d'évaluation le permet, il est acceptable que des thèmes plus ou moins reliés soient évalués dans toutes les épreuves. Cette possibilité s'inscrit dans la droite ligne du modèle éducatif de l'IB qui est contre tout cloisonnement des matières et des options.

Figure 27

Qu'est-ce qu'une bonne ou une mauvaise prévisibilité ?

Bonne prévisibilité	Mauvaise prévisibilité
Si l'on faisait l'inventaire de toutes les questions que l'on pourrait poser sur un thème, sur une option ou sur un texte en fonction de tous les mots-consignes, les questions posées par l'IB dans la matière en question y figureraient.	Les enseignants préparent excessivement les candidats à certaines questions lorsqu'ils remarquent qu'elles varient très peu d'une session à l'autre. Cette pratique pose problème pour les questions à réponse plus développée, qui imposent de montrer toute une gamme de compétences (connaissances, analyse et évaluation).
Les questions sont réutilisées de manière imprévisible.	Les questions possibles sont limitées en raison de mauvaises décisions prises lors de la conception de l'évaluation (par exemple, un texte prescrit choisi pour un cours ne comporte qu'un seul thème principal sur lequel il est possible de poser des questions). Cela conduit à une mauvaise prévisibilité due à la conception.

Une bonne évaluation s'appuie sur tout un éventail de tâches d'évaluation

Une question à choix multiple, une question à réponse brève, une question à réponse développée, une dissertation, un projet, un travail tiré d'un dossier et un travail de recherche constituent tous des exemples de l'éventail de tâches d'évaluation.

La composante d'évaluation ou l'outil d'évaluation est constitué d'une ou plusieurs tâches groupées pour assurer la continuité du thème ou du contenu, ou encore par commodité. Les outils d'évaluation ou composantes d'évaluation peuvent prendre la forme d'une épreuve d'examen, d'un dossier de travail, d'un projet ou d'un travail de recherche. Les concepts de tâche d'évaluation et de composante se chevauchent. Le candidat n'effectue parfois qu'une seule tâche parmi un certain nombre de tâches qui lui sont proposées dans une composante.

Plusieurs raisons expliquent que l'IB utilise une grande variété de types de tâches et de composantes d'évaluation. Tout d'abord, d'un point de vue historique et pratique, Peterson (2003) a dit au sujet de l'élaboration initiale de l'évaluation dans le Programme du diplôme que l'on avait à la fois une obligation et une occasion de tenir compte des différentes techniques d'évaluation utilisées dans les pays dans lesquels étaient situés les établissements scolaires où les candidats de l'IB essayaient principalement de s'inscrire, et ce principe peut être étendu au POP et au PEI. Il existe également des aspects de la validité (liés à l'adéquation avec l'objectif visé) qui requièrent une approche différente de l'évaluation. Enfin, l'utilisation de diverses techniques d'évaluation aide à réduire les possibilités d'iniquité dans le domaine de l'évaluation (voir également Linn, 1992 ; Brown, 2002). L'éventail des composantes du modèle d'évaluation pour une matière donnée et l'ensemble des tâches inhérentes garantissent une représentation adéquate des accomplissements du candidat par rapport à tous les objectifs définis pour cette matière.

Le rôle de l'évaluation se déroulant en classe et de l'évaluation interne

L'évaluation se déroulant en classe offre différentes possibilités d'évaluer les candidats dans des domaines qui ne sont pas adaptés aux examens. Son principal avantage est qu'elle permet d'engager les candidats dans une tâche plus longue, grâce à laquelle ils peuvent étudier un problème et montrer le développement de leur raisonnement sans les contraintes de temps caractéristiques des examens. Un large éventail de tâches d'évaluation peuvent ainsi uniquement être utilisées pour des évaluations se déroulant en classe.

Parmi les tâches qui se prêtent parfaitement à une évaluation se déroulant en classe, il est possible de citer les projets, le travail de terrain, les travaux pratiques en laboratoire et la recherche mathématique. Les épreuves orales, dans lesquelles un enseignant pose des questions et réagit aux propos du candidat, nécessitent également d'être organisées en classe, même si cela est susceptible de changer à l'avenir grâce à l'évolution de la technologie informatique.

Les travaux soumis à une évaluation interne ont d'autres avantages dans le contexte d'un diplôme reconnu internationalement. Ils permettent une grande flexibilité au niveau du choix des thèmes, tout en continuant de requérir un ensemble commun de compétences. Cela permet aux établissements scolaires de placer l'étude dans un contexte local, culturel ou géographique, ou d'établir des liens plus étroits entre la classe et le monde extérieur. Les écoles internationales, dont les élèves ont souvent un bagage culturel différent de celui des habitants du pays où l'établissement est implanté, peuvent utiliser les travaux soumis à l'évaluation interne pour mettre en place une participation plus importante dans la société ou l'environnement local. L'évaluation interne peut également être utilisée pour établir des liens avec des cultures éloignées, généralement en correspondant avec des établissements scolaires situés dans d'autres parties du monde. Brown (2002) a également montré la valeur de l'évaluation interne comme moyen d'introduire une diversité culturelle au sein de l'évaluation dans le Programme du diplôme. Cela encourage une plus grande vision internationale, en permettant l'utilisation d'un grand nombre d'approches culturelles et en donnant à chaque élève l'occasion de découvrir un éventail de valeurs culturelles.

En outre, l'évaluation interne donne souvent aux candidats la possibilité de choisir leur propre sujet ou thème selon leurs propres centres d'intérêt et leur donne une plus grande maîtrise de leur apprentissage. De par sa flexibilité, l'évaluation interne apporte une contribution précieuse à l'éducation des candidats ; elle renforce non seulement la validité du processus d'évaluation mais également celle de l'ensemble de l'apprentissage.

L'évaluation se déroulant en classe pose elle aussi des difficultés. Une des deux principales difficultés est qu'il est beaucoup plus compliqué de s'assurer qu'un candidat n'adopte pas une mauvaise conduite (par exemple, en demandant à quelqu'un de faire le travail à sa place). Et la disponibilité d'Internet ne fait qu'aggraver la situation. Les établissements scolaires sont les mieux placés pour reconnaître les cas où les candidats remettent un travail qu'ils n'ont pas réalisé eux-mêmes. Pour les aider à résoudre ce problème, les membres de la direction trouveront différentes ressources sur le site Web de l'IB.

La seconde difficulté majeure de l'évaluation se déroulant en classe est la charge importante qu'elle peut faire peser à la fois sur l'enseignant et sur le candidat. En effet, pour organiser des évaluations en classe, il faut rogner sur les heures d'enseignement, et les tâches données en interne sont généralement substantielles et requièrent des candidats qu'ils y consacrent beaucoup de temps. Bien qu'il soit approprié que les enseignants passent beaucoup de temps à familiariser les candidats avec les compétences et les processus requis pour l'évaluation interne, la tentation peut être grande pour les candidats et les enseignants de s'entraîner plus que nécessaire à effectuer la tâche donnée, et ce, afin d'obtenir le meilleur résultat possible, ce qui limite donc encore davantage le temps consacré à l'enseignement.

L'évaluation se déroulant en classe peut être notée soit en externe, soit en interne. En cas d'évaluation externe, le travail produit par le candidat est envoyé à l'IB et noté par un examinateur. La qualité de la décision de l'examineur peut ensuite être contrôlée de la façon habituelle (voir la section « [Le modèle d'assurance de la qualité](#) »). Les évaluations se déroulant en classe consistent généralement en de longs travaux qui sont difficiles à noter.

En cas d'évaluation interne, l'établissement scolaire (ou plus exactement l'enseignant) note le travail du candidat, puis l'IB vérifie que l'enseignant a correctement appliqué la norme globale en utilisant un processus appelé « révision de notation ».

Les avis sur l'évaluation interne sont partagés à travers le monde. Certains systèmes éducatifs attachent une grande importance au fait que les enseignants sont les plus à même de donner une opinion globale sur l'accomplissement du candidat, sans se limiter au seul « aperçu » permis par une évaluation unique. D'autres se basent sur les résultats des candidats pour déterminer la réussite des enseignants, ce qui incite donc fortement les enseignants à attribuer des notes élevées à toutes les évaluations internes.

Le jugement de l'enseignant peut aussi être altéré par sa connaissance des travaux antérieurs du candidat, car ils peuvent amener l'enseignant à avoir certaines attentes. Les enseignants ne savent parfois pas très bien quelles sont les limites de leur rôle lorsqu'ils guident et aident les candidats lors de la réalisation des travaux d'évaluation interne, et ils n'ont souvent qu'une vue restreinte des niveaux d'accomplissement que l'on trouve à l'échelle mondiale dans leur discipline. Lorsqu'ils évaluent les travaux de leurs propres candidats, les enseignants peuvent être grandement influencés par les niveaux d'accomplissement généraux existant dans leur établissement. Même si rien ne pousse à noter de manière bienveillante le travail d'un candidat, les enseignants auront du mal, du fait de la relation professionnelle établie avec les apprenants, à prendre des décisions objectives. Les études menées sur ce biais inconscient indiquent qu'il peut avoir des effets perceptibles.

Pour l'IB, l'utilisation de tâches d'évaluation riches de sens, passant obligatoirement par un travail réalisé en classe, et évaluées par une personne ayant vu le développement s'opérer chez les candidats, comporte des avantages plus grands que les risques présentés par l'évaluation interne. Par conséquent, les tâches d'évaluation interne font généralement partie intégrante de toute série d'évaluations de l'IB.

Pour terminer, les tâches se déroulant en classe peuvent également donner l'occasion aux candidats de montrer certains aspects du système éducatif de l'IB qui ne sont pas évalués.

Le travail en collaboration et les notes individuelles

Des difficultés se présentent dans un autre domaine : comment évaluer les tâches réalisées en collaboration, dans lesquelles les candidats doivent effectuer un travail collectif pour réaliser une activité ? Il s'agit d'un aspect important que les candidats rencontreront dans le milieu du travail, mais qui est souvent négligé dans les évaluations scolaires.

Dans certains cas, il est possible de déterminer facilement le rôle de chacun dans une activité de groupe. Par exemple, une représentation de danse permet de distinguer les compétences de chaque danseur et, par conséquent, de procéder à une notation individuelle. Dans ce type de situation, le travail de groupe n'est pas sujet à controverse.

Dans d'autres cas, il est impossible de savoir qui est responsable de quels aspects du travail final. Dans ce cas de figure, il est possible d'attribuer une note unique à l'ensemble du groupe, mais cette pratique ne tient pas compte des différences de contribution de chaque candidat au résultat global. Ainsi, même si un candidat réalise l'essentiel du travail et apporte la plus grande part d'expertise, il ne bénéficie d'aucune reconnaissance particulière. Étant donné que les résultats de l'IB sont utilisés à des fins de sélection, il semble que ce traitement ne soit pas juste ; c'est pourquoi les évaluations de groupe sont généralement à éviter lorsqu'elles ne permettent pas de juger l'accomplissement de chaque candidat.

Cette approche peut provoquer un biais culturel, car l'idée d'individualisme relève traditionnellement d'une idéologie propre à l'Europe occidentale. L'IB a bien conscience de ce problème, mais il fait la part des choses en s'appuyant sur l'idée que même dans les autres cultures, les résultats de l'évaluation sont souvent utilisés à des fins de sélection.

Une bonne évaluation tient compte des compétences plus larges du candidat et des compétences de pensée de haut niveau

Le système éducatif de l'IB vise bien au-delà du simple apprentissage de « faits » par les élèves. Cet objectif de longue date se retrouve dans la réflexion actuelle d'un certain nombre de gouvernements sur le besoin

de doter les élèves des compétences nécessaires au XXI^e siècle et des aptitudes recherchées dans le monde professionnel, ou dans d'autres initiatives de ce genre.

La philosophie de l'IB et son approche en matière de compétences des élèves se concentrent sur le profil de l'apprenant et sur son lien avec la sensibilité internationale. Ces éléments seront expliqués de façon plus détaillée dans la section « [Le socle commun à tous les programmes](#) ».

Une bonne évaluation tient compte de toute la gamme de résultats qu'un cours souhaite obtenir et permet aux candidats de montrer leurs compétences dans chaque domaine. Cependant, il est parfois souhaitable, voire uniquement faisable, de mesurer une petite partie de ces résultats, et une évaluation de qualité tient compte de ces limites. Pour l'IB, le meilleur système consiste à classer ces résultats en fonction des compétences de pensée de haut niveau, des compétences plus larges des candidats et de leur lien avec la sensibilité internationale.

Les compétences cognitives de haut niveau

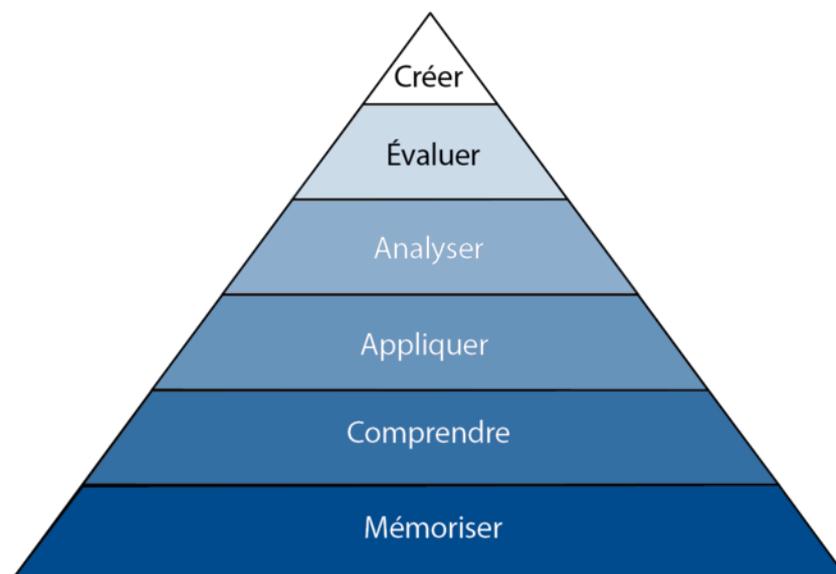
Les évaluations de l'IB cherchent à mesurer les compétences de pensée de haut niveau que sont la capacité d'évaluation et la capacité d'analyse, et non la simple aptitude à se rappeler des connaissances.

Cette remarque a été faite par Alec Peterson (2003), qui a donné forme à la philosophie pédagogique de l'IB. Alec Peterson a déclaré que ce qui importe n'est pas l'acquisition et la régurgitation de faits ou d'interprétations prémâchées de faits, mais le développement de facultés de l'esprit ou de modes de pensée qui peuvent être appliqués à de nouvelles situations et à de nouvelles présentations des faits au fur et à mesure qu'ils surviennent. Lors de la conférence des directeurs d'école du monde de l'IB, organisée à Singapour en 2011, Sugata Mitra a poussé l'argumentation plus loin. Il y a soutenu que, puisqu'Internet permettait aujourd'hui d'avoir une profusion de connaissances à disposition immédiatement, les connaissances n'avaient plus vraiment de valeur intrinsèque, et que c'était donc la capacité à analyser, à interpréter et à sélectionner les connaissances qui était exigée des citoyens du XXI^e siècle (Mitra, 2011).

Depuis longtemps, les évaluations de l'IB essaient de prêter une attention particulière aux compétences cognitives dites « supérieures » (Bloom *et al.*, 1956 ; Anderson et Krathwohl, 2001). Il peut y avoir des désaccords quant à la nature hiérarchique des niveaux proposés par Bloom ou quant au nombre de niveaux, mais sa taxonomie des objectifs pédagogiques fournit toutefois une structure utile qui permet d'exprimer la diversité des compétences requises. Les compétences supérieures décrites par Bloom exigent sans aucun doute un type d'évaluation différent. Les compétences des élèves en matière d'analyse, de synthèse et d'évaluation ne peuvent être correctement évaluées que si on leur demande d'analyser, de synthétiser et d'évaluer dans le détail. La seule façon réaliste de mesurer l'accomplissement de l'élève dans ces domaines est de se pencher sur ses réalisations et, du fait que les résultats d'une telle activité ne peuvent être rigoureusement établis, de telles évaluations doivent être relativement non structurées et ouvertes lorsque de nombreuses réponses correctes différentes sont possibles.

Figure 28

La taxonomie de Bloom : une manière de décrire une hiérarchie dans les compétences de pensée



Pour être pertinentes, les évaluations de l'IB doivent impérativement pouvoir reconnaître et récompenser l'accomplissement de l'élève dans ces compétences, en dépit des difficultés que cela pose en ce qui concerne la fiabilité et d'autres aspects de la validité. Les tests qui ne récompensent que le rappel de connaissances, de concepts et de techniques courantes ne sont pas adaptés aux objectifs du système éducatif de l'IB.

Les compétences de l'élève et le profil de l'apprenant

Aujourd'hui, l'enseignement s'apparente bien plus à un mode de pensée fondé sur la créativité, le raisonnement critique, la résolution de problèmes et la prise de décision. Il tient également à des méthodes de travail, notamment la communication et la collaboration, et à des outils de travail, y compris la capacité de reconnaître et d'exploiter le potentiel des nouvelles technologies, mais aussi d'éviter leurs écueils. Enfin et surtout, l'enseignement concerne la capacité de vivre en citoyen actif et engagé dans un monde protéiforme. À leur tour, ces citoyens exercent une influence sur ce qu'ils veulent apprendre et sur la façon d'apprendre, d'où une évolution du rôle des éducateurs.

(Schleicher, 2016)

D'aucuns prétendent de plus en plus que ce siècle requiert des compétences radicalement différentes de celles attendues des générations précédentes. Même si certains soutiennent que l'approche de la recherche qui sous-tend ces compétences du XXI^e siècle est en réalité valorisée depuis Socrate, la majorité s'accorde à dire qu'il est important que les élèves possèdent un large éventail de qualités pour être bien préparés à la vie qui les attend (voir, par exemple, les arguments avancés par Llewellyn *et al.*, 2014).

Un cadre des compétences du XXI^e siècle a notamment été défini par l'OCDE, RAND Education et le National Research Council (NRC), pour ne citer qu'eux. L'IB, quant à lui, décrit ces compétences dans son profil de l'apprenant.

Figure 29

Profil de l'apprenant de l'IB



Tous les aspects du profil de l'apprenant ne peuvent pas être évalués au moyen d'une évaluation sommative, mais plusieurs sont englobés dans la notion de compétences de pensée de haut niveau. Une bonne évaluation reconnaît l'importance de ces caractéristiques et, même si elle ne vise pas à mesurer ces compétences, elle peut donner aux élèves l'occasion de les développer. Il pourrait s'agir, par exemple, d'encourager des approches éthiques (intègres) des enquêtes et des expériences, en facilitant une révision par les pairs appropriée et en présentant des contextes inattendus aux élèves.

Pour de plus amples informations sur l'approche globale adoptée par l'IB en matière de compétences des élèves, veuillez vous reporter aux documents disponibles sur son [site Web](#) ou à la publication *Des principes à la pratique* spécifique à chaque programme.

La sensibilité internationale et la compréhension interculturelle

Les programmes de l'IB sont étudiés par des élèves situés dans de nombreux pays et représentant toute une palette de nationalités. L'IB entend les amener non seulement à atteindre les objectifs pédagogiques de ses programmes, mais aussi à développer « la sensibilité nécessaire pour contribuer à bâtir un monde meilleur et plus paisible, dans un esprit d'entente mutuelle et de respect interculturel » et à « comprendre que les autres, en étant différents, puissent aussi être dans le vrai » (extraits de la déclaration de mission de l'IB, 2002). L'enseignement prôné par l'IB se caractérise donc à la fois par un contexte international et par un objectif de compréhension interculturelle, deux éléments qui doivent se retrouver dans l'évaluation.

Pour y parvenir, le plus important est de faire appel à des spécialistes de l'éducation, notamment des auteurs d'épreuves d'examen et des personnes chargées du développement des programmes d'études, issus de contextes culturels très divers. La mission de l'IB tient à ne pas occulter les différences, mais au contraire à les aborder de façon à ce que les élèves les explorent sans se trouver défavorisés.

Dans certaines disciplines, il est possible de favoriser la diversité culturelle, en tenant compte des différentes perspectives culturelles dans le programme d'études. La biologie, la chimie, la psychologie et les arts visuels offrent des exemples de cette approche. Dans les trois premiers cours, les options permettent aux établissements scolaires de choisir un contenu qui respectera, dans une certaine mesure, leurs traditions culturelles concernant l'enseignement de la matière.

Dans d'autres disciplines, la sensibilité internationale est encouragée au moyen du matériel et des sources d'inspiration que l'élève est incité à utiliser. C'est notamment le cas des matières artistiques et des cours de langue et littérature, mais cette notion peut également intervenir dans un éventail plus large de tâches d'évaluation interne.

La sensibilité internationale ne se réduit pas au fait de connaître et de comprendre d'autres cultures. Les attitudes affichées et les actions entreprises en sont d'autres aspects fondamentaux. Il est difficile d'évaluer les attitudes à l'aide des évaluations scolaires traditionnelles, qui se concentrent davantage sur les accomplissements que sur les qualités affectives.

L'IB parvient à le faire par l'intermédiaire des éléments non évalués de ses programmes, comme le programme créativité, activité, service dans le Programme du diplôme et le projet communautaire dans le PEI. Les candidats ne peuvent se voir attribuer le diplôme de l'IB ou le certificat du PEI s'ils ne terminent pas avec succès ces composantes, qui, même si elles ne sont pas directement évaluées, influent grandement sur le résultat global du processus d'évaluation de l'IB.

Figure 30
Action fondée sur des principes



Le fait de permettre aux candidats de choisir les questions auxquelles ils souhaitent répondre peut être considéré comme la meilleure façon de répondre aux différentes exigences internationales en matière d'évaluation, mais cela pose certains problèmes du point de vue de la comparabilité entre les options. Cela est toujours le cas lorsque plusieurs questions sont proposées ou que les tâches d'évaluation sont très ouvertes. Même la définition d'une « difficulté égale » est difficile à donner lorsque les candidats sont issus de milieux scolaires très différents. En règle générale, il est plus facile de garantir la comparabilité avec des tâches communes qui permettent aux candidats de s'appuyer sur leurs expériences personnelles pour rédiger leurs réponses. Il revient alors à l'examineur de réussir à maintenir une norme commune, mais ce défi est moins grand que celui de noter selon la même norme deux tâches qui peuvent présenter des niveaux de difficulté différents.

Il est possible d'obtenir des renseignements sur la comparabilité en analysant les résultats des candidats. Cette analyse est examinée plus en détail dans la section « [L'attribution des notes finales \(et la totalisation\)](#) ».

L'évaluation effectuée dans un contexte international présente des défis supplémentaires (en matière d'équité), qui sont plus importants que ceux normalement rencontrés dans un système national. Des questions qui sont parfaitement appropriées dans un pays peuvent devenir complètement inappropriées dans un autre. Les questions portant sur les sports, les voyages, les divertissements, les événements historiques, et même sur la météo, doivent être préparées avec beaucoup de soin. Il semblerait que la seule façon de contourner ce problème soit de rédiger des questions d'examen dénuées de tout contexte socioculturel, à l'exception du plus petit dénominateur commun. Toutefois, cela rendrait non seulement les questions d'examen très limitées et ennuyeuses mais irait également à l'encontre de la philosophie de l'évaluation de l'IB et des bonnes pratiques de l'évaluation, en ce sens que cela ne permettrait pas de garantir la validité au moyen de tâches basées sur le contexte. Pour que l'apprentissage soit bon, il est vital que le travail et l'évaluation tiennent compte du contexte.

Il existe deux façons de résoudre ce dilemme. Premièrement, des informations sur le contexte peuvent être fournies aux candidats, grâce aux précisions apportées dans le contenu du programme de la matière, en leur donnant des études de cas sur lesquelles les questions sont basées, ou encore dans la question d'examen elle-même (à condition que ces informations ne soient pas trop longues et qu'elles ne distraient donc pas les candidats de l'objet de l'évaluation).

La seconde méthode consiste à utiliser des tâches d'évaluation et des questions d'examen plus ouvertes, qui permettent aux candidats de choisir leur propre contexte. Dans cette méthode, la notation doit se concentrer sur une compréhension plus profonde plutôt que sur une simple connaissance du contenu du programme de la matière, car il n'y aura aucune base commune en ce qui concerne le contenu. Cela s'accorde parfaitement avec les principes de l'IB en matière d'évaluation des candidats.

Figure 31

Gamme de normes culturelles et contextes



Quand bien même ces deux méthodes seraient utilisées, les candidats pourraient avoir à réaliser des tâches d'évaluation dont le contexte n'est pas familier dans leur milieu socioculturel. Une fois de plus, cela s'accorde avec la philosophie de l'IB en matière d'évaluation, en ce sens que l'un des objectifs des programmes est de rendre les élèves plus ouverts aux autres façons de faire les choses, plus conscients du monde et plus aptes à fonctionner dans un milieu culturel qui ne leur est pas familier. L'une des exigences concernant la pensée de haut niveau est d'être capable d'utiliser ses connaissances dans des situations inconnues. Il est donc parfaitement approprié d'inclure de tels éléments dans l'évaluation pourvu qu'ils aient un effet égal sur les élèves issus de différents milieux culturels.

Un nombre important d'élèves de l'IB passent les examens dans une langue qui n'est pas leur meilleure langue. Presque tous ces cas concernent l'anglais, car les élèves qui travaillent en français ou en espagnol (les deux autres langues principales utilisées pour les évaluations de l'IB) sont en général des locuteurs natifs. Il faut donc porter une grande attention à la façon dont sont formulées les questions de manière à ne

pas désavantager les élèves qui ne sont pas des locuteurs natifs. Ces précautions sont prises au moment de la révision des épreuves.

L'évaluation sommative de l'IB, tout comme la grande majorité des systèmes d'évaluation officiels, est très individualiste. Comme l'a souligné Brown (2002), cela est en grande partie dû au fait que le Programme du diplôme appartient à une tradition propre à l'Europe occidentale et que les sociétés qui la composent sont par nature individualistes. Les candidats sont presque uniquement évalués sur ce qu'ils accomplissent seuls. On peut considérer cela comme étant culturellement inéquitable, car, dans un certain nombre de cultures, la contribution de l'individu est toujours moins importante que celle du groupe ; dans ces cultures, c'est ce qu'accomplit le groupe qui est important. Il est également vrai qu'en matière d'équité individuelle, il y a des individus qui travaillent mieux en équipe que seuls, et vice-versa. De plus, il est courant, à la fois dans la salle de classe et dans le monde du travail, que les individus travaillent de manière interdépendante plutôt que de manière indépendante.

À quoi ressemble une bonne évaluation sur ordinateur ?

Il a été indiqué précédemment que l'évaluation sur ordinateur permettait des évaluations de meilleure qualité, en utilisant l'ordinateur pour poser des questions qui ne pourraient pas être posées dans les examens sur papier. Il s'en dégage trois éléments clairs qui définissent une bonne évaluation sur ordinateur.

- L'évaluation sur ordinateur s'appuie sur les possibilités offertes par cet outil de proposer des tâches plus valides, du fait de leur authenticité, de leur pertinence, de la réduction du biais ou de tout autre facteur.
- Elle lève les obstacles qui empêchent certains candidats de prendre part à l'évaluation ou améliore l'ensemble du processus, notamment en supprimant certains risques et les retards d'expédition.
- Elle n'impose pas de nouveaux obstacles (notamment une interface difficile à utiliser et des problèmes techniques) aux candidats qui passent l'évaluation.

À l'inverse, si certaines fonctionnalités (comme la vidéo) sont utilisées uniquement parce qu'elles sont disponibles et non pour accroître la validité de l'évaluation, alors c'est le signe d'une évaluation sur ordinateur de mauvaise qualité. La technologie peut et doit faciliter le processus d'évaluation, mais elle ne doit pas déterminer la direction qui lui est donnée.

Les principes de l'IB en matière d'évaluation

Les cinq points ci-dessous résument les principes qui sous-tendent les évaluations de l'IB.

Les évaluations de l'IB doivent :

1. être valides pour les buts qu'elles visent. En d'autres termes, elles doivent trouver un équilibre entre les exigences contradictoires posées par la pertinence de construit, la fiabilité, l'équité (c'est-à-dire l'absence de biais), la comparabilité avec d'autres évaluations et la capacité de gestion pour les candidats, les établissements scolaires et l'IB ;
2. avoir un effet de remous positif, c'est-à-dire que leur conception doit favoriser un enseignement et un apprentissage de qualité ;
3. convenir au plus grand nombre possible de candidats, en leur permettant de montrer le niveau qu'ils ont personnellement atteint ;
4. faire partie intégrante du programme de l'IB, et non être considérées comme des éléments isolés. Il convient de se demander si les évaluations favorisent la simultanéité des apprentissages et soutiennent l'expérience globale de l'apprenant ;
5. soutenir la mission de l'IB et renforcer les compétences des élèves, notamment les qualités suivantes : chercheurs, informés, sensés, communicatifs et sensibles à la réalité internationale.

Qu'entend-on par « pratique » ?

- Les pratiques de l'évaluation de l'IB englobent les évaluations sommatives organisées dans le cadre du Programme du diplôme, du POP et du PEI.
- Elles ne s'appliquent pas au PP, dans lequel l'IB n'organise aucune évaluation sommative, ni à l'évaluation formative organisée par l'établissement scolaire.

Cette section présente les pratiques employées par l'IB pour produire les résultats des candidats passant les évaluations qui sont notées ou révisées en externe. Elle n'aborde pas les évaluations qui ne sont pas de la responsabilité de l'IB, telles que les évaluations conçues par les enseignants pour être utilisées dans le PP.

Un principe indique *pourquoi* quelque chose est fait, et une pratique décrit *comment* cela est mené à bien. Cette section explique donc les pratiques de haut niveau auxquelles il est fait appel pour garantir la validité des résultats de l'évaluation.

Elle traite ensuite en détail les procédures, qui décrivent les différentes étapes à suivre pour appliquer chaque pratique. Lorsque ces procédures concernent les établissements scolaires, elles sont expliquées en détail dans le document *Procédures d'évaluation* spécifique à chaque programme.

Transmission des résultats des candidats

- L'objectif affiché par la déclaration de mission de l'IB est de former des jeunes capables de bâtir un monde meilleur, et une évaluation réussie devra donc soutenir cet objectif.
- Les notes finales de l'IB ont une signification, et les seuils d'attribution des notes finales sont établis de façon à en tenir compte.
- Même si les notes finales ne sont qu'un aperçu très simplifié des accomplissements des candidats, elles permettent aussi aux interlocuteurs tels que les universités, les employeurs ou les divers établissements scolaires de disposer d'une base pour décider de leur sélection.
- Si l'IB ne fournissait que des informations plus complexes et globales, d'autres auraient à simplifier ces preuves pour prendre des décisions éclairées en vue d'une sélection, ce qui serait certainement fait avec une attention moins aigüe que celle apportée par l'IB lors de l'établissement des seuils d'attribution des notes finales.
- Les accomplissements des candidats ne se réduisent pas aux résultats obtenus aux examens, et même en fonctionnant avec des notes finales, il ne faut pas perdre de vue qu'un résultat décevant pour un candidat pourra être un grand accomplissement pour un autre.

Le Baccalauréat international (IB) a pour but de développer chez les jeunes la curiosité intellectuelle, les connaissances et la sensibilité nécessaires pour contribuer à bâtir un monde meilleur et plus paisible, dans un esprit d'entente mutuelle et de respect interculturel.

(Extrait de la déclaration de mission de l'IB, 2002)

Qu'entend-on par « note finale de l'IB » ?

Les résultats d'un candidat à l'issue des évaluations de l'IB portent le nom de « notes finales ». Ces notes finales décrivent la qualité du travail démontré par le candidat dans ses réponses.

Chaque note finale est associée à un descripteur, qui est publié par l'IB. Étant donné que ces descripteurs reflètent la qualité du travail attendue des candidats en fonction de leur âge, ils sont différents pour le Programme du diplôme, le POP et le PEI.

Figure 32

Exemples de descripteurs des notes finales du Programme du diplôme

Descripteurs des notes finales du groupe 3 (Individus et sociétés)

Note finale 7

L'élève met en œuvre des compétences de pensée critique qui témoignent de sa sensibilité par rapport aux concepts, de sa perspicacité, de ses connaissances et de sa compréhension. Il s'avère extrêmement apte à apporter des réponses pleinement développées, structurées de manière logique et cohérente, et illustrées à l'aide d'exemples appropriés. Il emploie avec précision la terminologie spécifique à la matière et démontre une bonne connaissance de la documentation relative à la matière. Il est capable d'analyser et d'évaluer des preuves, ainsi que de synthétiser des connaissances et des concepts. Il a conscience des différents points de vue ainsi que des préjugés subjectifs et idéologiques. Il est capable de tirer des conclusions raisonnables, même si celles-ci sont parfois approximatives. Son travail fait systématiquement état d'une réflexion critique et approfondie. Il est capable d'analyser et d'évaluer des données ou de résoudre des problèmes de manière extrêmement compétente.

Note finale 6

L'élève fait preuve d'une connaissance et d'une compréhension précises. Ses réponses sont cohérentes, structurées de manière logique et bien développées. Il emploie systématiquement la terminologie appropriée. Il est capable d'analyser, d'évaluer et de synthétiser des connaissances et des concepts. Il connaît les recherches, les théories et les questions pertinentes et a conscience des différents points de vue et contextes dans lesquelles elles s'inscrivent. Son travail fait systématiquement état d'une réflexion critique. Il est capable d'analyser et d'évaluer des données ou de résoudre des problèmes de manière compétente.

Note finale 5

L'élève fait preuve d'une connaissance et d'une compréhension solides de la matière et a recours à la terminologie spécifique à la matière. Ses réponses sont cohérentes et structurées de manière logique mais ne sont pas pleinement développées. Il est capable d'apporter des réponses pertinentes en essayant d'intégrer des connaissances et des concepts. Il a tendance à produire un travail plus descriptif qu'évaluatif même s'il fait preuve d'une certaine capacité à présenter et à développer des points de vue opposés. Son travail fait parfois état d'une réflexion critique. Il est capable d'analyser et d'évaluer des données ou de résoudre des problèmes.

Les descripteurs généraux des notes finales doivent être identiques pour toutes les matières d'un programme, mais l'IB les place souvent dans des contextes spécifiques aux matières pour les rendre plus compréhensibles dans chaque matière. Il convient de comprendre que la qualité du travail ne change pas en fonction du contexte de la matière. Ainsi, une note finale de 4 en langue doit avoir la même signification qu'une note finale de 4 dans une matière scientifique ou artistique. Cette idée se retrouve dans l'approche adoptée par l'IB dans ses programmes, où toutes les notes finales sont d'importance équivalente, mais elle fait l'objet de nombreuses discussions parmi les spécialistes de l'éducation qui se posent la question suivante : comment peut-on comparer un accomplissement réalisé dans deux matières différentes ? Est-il même utile de faire une telle comparaison ? Ce concept est examiné de manière plus approfondie dans la section « Comparabilité ».

Figure 33

Comment ces deux travaux peuvent-ils être comparés ?



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

en prenant le logarithme népérien de chaque membre de l'équation

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

(b) les coordonnées de Q sont (1,0)

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

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

$$y=x-1$$

(c) soit A l'aire demandée

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

utilisez l'intégration par parties pour trouver $\int \ln x dx$

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

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

Dans le contexte des évaluations de l'IB, l'argument nous amène à la validité des objectifs visés par les notes finales. Ces notes finales ont pour but de permettre aux interlocuteurs de comparer les niveaux atteints par les candidats. Par conséquent, il est judicieux de suivre des méthodes statistiques et qualitatives pour parvenir à la parité des significations données aux notes finales.

Quelle est la différence entre les notes et les notes finales ?

Notes et notes finales sont deux choses différentes.

Figure 34

Ce qui importe, ce n'est pas uniquement la distance parcourue à pied, mais aussi la nature du sol foulé.



On s'attendrait à ce que les « bons » candidats aient terminé la plus grande partie de la tâche.

On s'attendrait à ce que les « bons » candidats aient terminé une partie de la tâche seulement.



De nombreuses métaphores permettent d'expliquer la différence entre les notes et les notes finales. Par exemple, dans les illustrations ci-dessus, la distance parcourue pourrait représenter les notes. Il s'agit d'une mesure courante du chemin parcouru par une personne. Mais pour comprendre dans quelle mesure il s'agit d'un accomplissement, il faut tenir compte de la nature du sol foulé. Et c'est cette idée qui est reprise dans la notion de notes finales.

- Une note est le reflet de la proportion de la tâche réalisée par le candidat.
- Une note finale tient compte de la difficulté de la tâche pour indiquer dans quelle mesure il faut être impressionné par la note obtenue par le candidat.

Étudiez les deux exemples qui suivent. Dans le premier cas, on s'attendrait à ce qu'un candidat de 16 ans réussisse presque toute la tâche pour avoir une « bonne » note finale, alors que dans le second cas, les attentes seraient bien moins élevées pour attribuer une « bonne » note finale.

Figure 35

Les exigences à l'égard des bons candidats seraient moins élevées en cas de tâche difficile qu'en cas de tâche facile.

The figure shows two musical excerpts side-by-side. On the left is the beginning of 'Three Blind Mice', featuring a simple melody in G major with lyrics: 'Three blind mice, Three blind mice, See how they run, See how they run. They all run af-ter the fat-ter's wife, who cut off their tails with a carving knife, Did you ev-er see such a sight in your life, as three blind mice?'. On the right is the beginning of 'SONATA XIV.' by Beethoven, marked 'Allegro (♩ = 64)'. It consists of multiple staves of complex piano music with various technical markings and dynamics.

On s'attendrait à ce que les « bons » candidats aient terminé la plus grande partie de la tâche.

On s'attendrait à ce que les « bons » candidats aient terminé une partie de la tâche seulement.

Ce raisonnement amène à l'un des défis posés par l'organisation d'une évaluation de qualité. Les candidats doivent avoir la possibilité de montrer leur plein potentiel, ce qui est souvent incompatible avec des tâches trop simples. Et inversement, si les difficultés sont trop grandes, il est probable que la moitié de la cohorte la moins brillante ne parviendra même pas à commencer la tâche, d'où une impossibilité d'utiliser les notes finales pour faire une distinction entre ces candidats.

L'autre risque présenté par l'utilisation de tâches trop simples est de commencer à mesurer l'exactitude plutôt que la compréhension. Les candidats ayant commis des erreurs mineures pourraient ne pas obtenir les meilleures notes finales, quand bien même ils auraient une bonne compréhension des thèmes, et ce n'est pas ce que vise l'évaluation.

La tyrannie des notes finales : un moindre mal

Réfléchissez aux connaissances, aux compétences et à l'expérience qui font l'excellence d'un chef. Ses connaissances portent sur les combinaisons d'ingrédients et de saveurs qui lui permettront de composer un plat parfaitement équilibré. Ses compétences interviennent pour le choix et la préparation des ingrédients, la détermination du mode de cuisson (plaque chauffante ou four) et la présentation. Et son expérience concerne l'utilisation des techniques pour atteindre la perfection, l'appréciation d'un plat merveilleusement cuisiné et la présentation d'un ensemble d'ingrédients et de plats pour procurer une immense satisfaction à table.

Figure 20

Connaissances, compétences et expérience

Réduisez maintenant toute cette complexité en une note finale unique sur 7 pour déterminer qui est le meilleur chef. Le résultat est presque vide de sens. Serait-il utile d'augmenter l'échelle et de choisir une note finale sur 100, voire sur 1 000 ? La réponse sera probablement « non ». Cela permettrait un plus grand champ de différenciation, mais le problème fondamental, à savoir comparer des ensembles de compétences différents, ne serait pas résolu.

Ce problème est précisément celui que pose l'évaluation : comment représenter la complexité des connaissances, de la compréhension et des capacités des apprenants en un résultat unique qui en ferait la synthèse ? Même si toutes les informations sur un candidat pouvaient être recueillies, cela ne signifierait pas pour autant qu'il serait possible de lui attribuer une note finale reflétant parfaitement ses talents.

Une autre possibilité, souvent suggérée, serait de ne pas attribuer de note finale du tout, mais peut-être de fournir pour chaque candidat une description personnalisée de ce qui a été bien réussi et des aspects à renforcer. Une telle approche respecte les principes d'un enseignement et d'un apprentissage de qualité, mais elle présente l'inconvénient majeur (que certains de ses partisans considèrent néanmoins comme une force) de compliquer considérablement les comparaisons entre candidats.

Pour tenter de résoudre ce problème, il est utile de revenir au but des évaluations de l'IB, qui est d'aider les candidats à pouvoir poursuivre leurs études ou à intégrer le monde du travail. Cela signifie qu'une forme de sélection devra être pratiquée par les établissements d'accueil (généralement les universités pour le Programme du diplôme ou le POP et les établissements scolaires pour le PEI) afin de choisir les candidats qui seront acceptés.

À l'évidence, si les autres critères sont moins fiables que les examens, compter davantage sur ces critères signifiera faire reposer la sélection sur des décisions moins fiables.

(Cresswell, 1986, p. 37 – 54)

Cette citation figure dans l'introduction du présent document, mais elle mérite d'être rappelée. Si une sélection doit avoir lieu, alors l'IB est tenu d'apporter son soutien aux élèves en rendant ces décisions en matière de sélection aussi justes et pertinentes que possible. Si les établissements d'accueil ne se voient remettre que des bilans descriptifs pour chaque élève, ils auront à trouver une autre méthode de comparaison, qui sera certainement moins fiable et uniforme que les résultats aux examens exprimés sous forme de notes finales. Réfléchissez, par exemple, à la validité d'un court entretien avec un tuteur et à tous les facteurs susceptibles d'influencer le résultat qui ne devraient pas servir de base pour la sélection.

L'intention n'est pas de laisser entendre que les évaluations sommatives sont un moyen idéal ou particulièrement bon pour procéder à une telle sélection, mais elles sont plus justes que les autres possibilités. Plus important encore, l'IB s'efforce constamment de rendre cette méthode la plus juste, utile et fiable possible.

Figure 21

Une distinction devrait-elle être faite entre les deux candidats de chaque groupe ?



Dans l'exemple ci-dessus, la différence entre les deux candidats du premier groupe est celle qui est faite entre les notes finales de 5 et 6 dans les descripteurs généraux des notes finales du Programme du diplôme. Si vous deviez choisir entre ces deux candidats, votre décision serait probablement juste, et ces candidats obtiendraient sans doute le même résultat s'ils devaient repasser le test. Dans le second groupe, la différence risque de ne pas être significative, et les deux candidats ont probablement la même aptitude.

Cet exemple prouve que toutes les différences ne sont pas significatives et que l'utilisation des notes finales donne une indication des aspects qui permettent de distinguer deux candidats. Il est certainement vrai que pour un candidat se situant à la limite entre deux notes finales (que ce soit juste au-dessus ou juste au-dessous du seuil d'attribution) n'importe laquelle des deux notes finales est juste, mais, pour la plupart des candidats, une différence entre les notes finales représente un écart d'accomplissement important.

Ce raisonnement amène à un débat sur le nombre de notes finales à employer. Avec seulement deux notes finales (réussite ou échec), la plupart des candidats auraient une note juste, mais les conséquences seraient sérieuses pour ceux qui se situent à la limite. À l'inverse, avec 20 notes finales, bien plus de candidats se situeraient à une limite, mais, pour chacun, les conséquences d'un rattachement à la mauvaise catégorie seraient moins graves. Ce concept a fait l'objet d'une étude plus poussée par Cresswell (1986).

En général, l'IB a retenu sept notes finales pour représenter le nombre de catégories importantes découlant de ses évaluations et pour trouver un juste équilibre entre le nombre de candidats se situant à une limite et les conséquences de l'attribution d'une note finale inappropriée.

Qu'est-ce qu'une session d'examens réussie ?

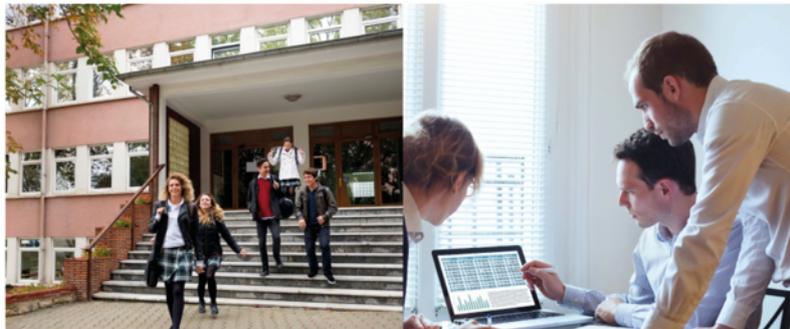
Figure 36

La réussite d'une session dépend du point de vue adopté.



Je connaissais les réponses à toutes les questions.

Nous avons pu sélectionner les bons élèves pour nos cours.



Nos candidats ont mieux réussi que ceux des établissements voisins.

Aucune erreur n'a été commise dans les questions et dans la notation.

La forme donnée à la réussite dépend essentiellement du point de vue adopté, et aucun ne vaut plus qu'un autre. L'IB met tout autant l'accent sur le haut niveau de validité que sur la facilité d'utilisation.

- L'évaluation doit donner à chaque candidat des chances égales de montrer ses capacités.
- L'expérience doit être la plus simple possible pour les établissements scolaires comme pour les candidats.
- Tous les interlocuteurs doivent garder confiance dans les résultats (notes finales) publiés.

L'accomplissement ne se réduit pas aux notes finales

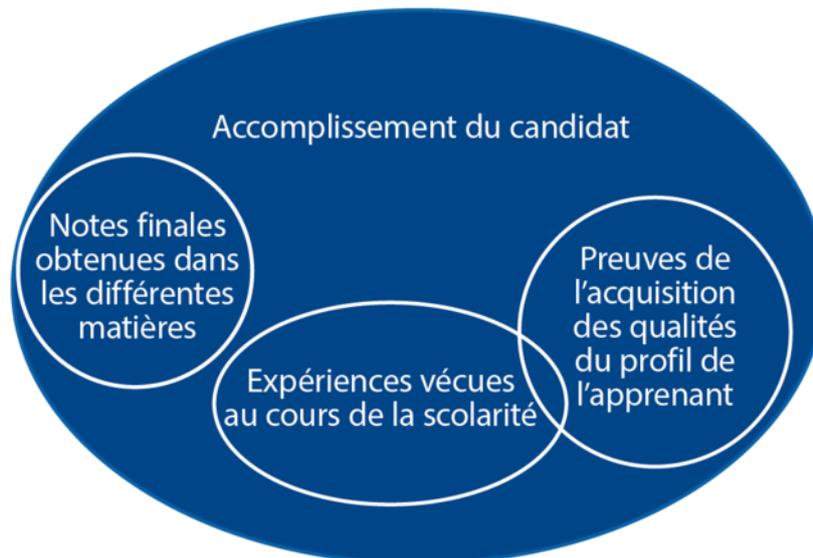
Comme l'indique la déclaration de mission de l'IB, le système éducatif de l'IB vise bien plus que la simple attribution d'une série de notes finales. Cela transparaît dans le profil de l'apprenant et le but de l'IB est expliqué dans la publication *En quoi consiste le système éducatif de l'IB ?*.

Les résultats de l'évaluation ne peuvent porter que sur un nombre restreint d'aspects de la déclaration de mission. Compte tenu du système de compensation utilisé dans les examens de l'IB, quel pourcentage des points attribués dans l'évaluation de mathématiques devrait être associé à la qualité « altruiste » ou « audacieux » (en supposant qu'il soit raisonnable d'attribuer une « valeur numérique » à ces qualités

essentielles) ? Les évaluations de l'IB reposent cependant sur le principe qu'elles doivent avoir un effet de remous positif qui facilitera l'atteinte des objectifs de l'IB.

Figure 37

L'accomplissement d'un candidat va bien au-delà de ce qui peut être prouvé.



Lorsque l'on rend compte des résultats des évaluations d'un programme de l'IB, il est important de ne pas s'arrêter aux notes finales pour illustrer l'ampleur de l'accomplissement d'un candidat.

L'IB ne consigne que le résultat final du candidat, sans donner aucune indication sur les difficultés qu'il a rencontrées pour parvenir à ce résultat ou pour atteindre son plein potentiel. L'IB prend toute la mesure de ces deux éléments, mais il estime qu'il ne peut les mesurer efficacement avec l'évaluation sommative. Une telle évaluation relève donc de la responsabilité de l'établissement scolaire, car il a une vision globale du travail du candidat.

Même si certaines méthodes permettent de calculer la « valeur ajoutée » de l'accomplissement d'un candidat ou les « notes finales prévues » en se basant sur l'accomplissement antérieur, l'IB recommande de faire preuve de prudence si elles sont utilisées pour avoir une idée de la réussite d'un candidat. Ces mesures reposent en effet sur le profil d'un élève moyen, alors que chaque élève présente un mélange unique de caractéristiques et de qualités personnelles, qui devraient être convenablement mises en valeur par les personnes ayant eu l'occasion d'apprendre à ses côtés pendant toute la durée d'un programme.

Processus d'évaluation – Rôles et responsabilités

- Chaque rôle occupé dans le processus d'évaluation est assorti de responsabilités particulières et d'un ensemble de compétences.
- Il peut arriver que les mêmes personnes remplissent toutes ces fonctions à différentes étapes du cycle d'évaluation.
- Certaines responsabilités incombent à l'IB et d'autres sont conférées à des spécialistes de la communauté de l'IB. Dans le dernier cas de figure, il incombe à l'IB de procéder à la validation finale et de veiller au maintien de la qualité de tous les aspects des évaluations.

Figure 38

Les responsabilités des principaux acteurs du cycle d'évaluation

<p>Examineur principal</p> <ul style="list-style-type: none"> • Il est responsable d'une composante donnée. • Il prend la décision finale concernant la note à attribuer aux réponses des candidats dans cette composante. • Il s'assure que tous les examinateurs comprennent la norme de notation définie pour la composante. • Il aide l'examineur en chef à établir les seuils d'attribution des notes finales pour la composante. 	<p>Examineur en chef</p> <ul style="list-style-type: none"> • Il a une vue d'ensemble de toutes les composantes de sa matière (ou de son groupe de matières). • Il veille à la cohérence des normes entre les différentes composantes, y compris pour la rédaction des épreuves. • Il tranche tous les débats portant sur l'évaluation. • Il recommande des seuils d'attribution des notes finales à l'IB.
<p>Examineur</p> <ul style="list-style-type: none"> • Il note le travail des candidats en respectant la norme établie par l'examineur principal. 	<p>IB</p> <ul style="list-style-type: none"> • Il est responsable de tous les aspects de l'évaluation. • Il est responsable des processus liés à l'évaluation, tels que le recrutement des examinateurs, le contrôle de la qualité du travail des examinateurs et la publication des résultats. • Il prend des décisions concernant les cas de mauvaise conduite ou de mauvaise administration, mais aussi des dispositions spéciales et des mesures compensatoires. • Il accepte ou conteste les seuils d'attribution des notes finales recommandés par l'examineur en chef.

L'examineur principal et l'examineur en chef

L'examineur principal est chargé d'une composante (par exemple, l'évaluation interne ou l'épreuve 2) et ses principales responsabilités sont les suivantes :

- il décide des réponses qui doivent recevoir des points (et établit la norme de notation) ;

- il explique cette norme aux examinateurs qui composent son équipe et révisé le barème de notation et toutes les directives supplémentaires destinées aux examinateurs, y compris les copies d'entraînement ;
- il aide l'examineur en chef à établir les seuils d'attribution des notes finales pour sa composante.

En bref, l'examineur principal est la personne qui doit gérer tous les sujets de préoccupation liés à la notation d'une épreuve donnée. Il bénéficie du soutien de plusieurs autres examinateurs expérimentés (les examinateurs superviseurs) qui débattent des problèmes pour lui faire des recommandations.

Il est exceptionnel que l'examineur principal d'un examen ne fasse pas partie de l'équipe de rédaction de l'épreuve, mais il ne s'agit pas d'une condition obligatoire pour occuper ce rôle.

L'examineur principal est engagé par l'IB pour une session d'examens donnée, mais n'est pas membre de son personnel. Pour plus de cohérence, il est généralement fait appel au même examinateur principal durant plusieurs sessions.

L'examineur en chef est quant à lui responsable du maintien de la qualité de plusieurs composantes connexes. Dans le Programme du diplôme et le POP, cela signifie qu'il est chargé de toute une matière, alors que, dans le cadre du PEI (qui ne compte qu'une composante par discipline), il est chargé d'un groupe de matières (par exemple, Sciences ou Individus et sociétés). Il joue le rôle de spécialiste de l'éducation dans son domaine et règle tout différend né entre les examinateurs principaux.

L'examineur en chef est chargé de :

- veiller à l'adéquation des normes dans chaque composante et entre les composantes ;
- garantir la cohérence entre les composantes lors de la notation et de la rédaction des épreuves ;
- régler tout désaccord entre ou avec les examinateurs principaux ;
- diriger le processus d'attribution des notes finales et recommander des seuils d'attribution des notes finales à l'IB ;
- jouer le rôle d'ambassadeur pour le compte de l'IB.

Les examinateurs en chef sont également invités à collaborer avec l'IB pour discuter de ses évaluations et les améliorer. La plupart des examinateurs en chef sont aussi examinateurs principaux pour les composantes dont ils sont responsables et bénéficient ainsi d'une expérience directe des réponses des candidats dans la discipline concernée. Contrairement aux examinateurs principaux, les examinateurs en chef doivent aussi veiller à la qualité des épreuves lors de la phase de rédaction.

Les examinateurs en chef ne font pas partie du personnel de l'IB, mais ils sont sous contrat avec l'organisation pour une période allant de deux à sept ans. Cette limite est fixée afin d'encourager la progression et de veiller à ce que les évaluations de l'IB ne soient jamais éculées.

Dans le cas des matières présentées par un petit nombre de candidats, l'IB n'a pas d'intérêt à nommer un examinateur en chef. Le cas échéant, il fera appel à un « examinateur responsable » qui se chargera de veiller à la cohérence entre les composantes et de lui recommander des seuils d'attribution des notes finales. Cette personne occupera également la fonction d'examineur principal pour une composante.

En règle générale, les examinateurs principaux et les examinateurs en chef participent également au cycle de révision du programme (qui n'est pas abordé dans le présent document).

Les autres examinateurs

Les examinateurs ont pour responsabilité de noter le travail des candidats en respectant la norme établie par l'examineur principal. Ils doivent prouver qu'ils ont compris cette norme et qu'ils l'appliquent au moyen du modèle d'assurance de la qualité (voir la section « [Le modèle d'assurance de la qualité](#) »).

Les examinateurs qui souhaitent corriger les travaux d'une session proposent leur candidature à l'IB. Ils doivent être spécialisés dans la matière visée (qu'ils auront généralement enseignée) et avoir de l'expérience dans l'enseignement auprès d'élèves de la tranche d'âge concernée. L'IB vérifie que les candidats remplissent bien ces conditions auprès des personnes désignées pour fournir une recommandation. En règle générale, les examinateurs peuvent noter uniquement les travaux d'une composante d'évaluation interne (travaux réalisés dans le cadre des cours) et d'une composante de l'examen. Les périodes de notation de ces deux composantes ne se chevaucheront pas. Les examinateurs

sont rémunérés pour les copies d'examen notées « en conditions réelles ». Les copies de validation et de contrôle (voir la section « [Le modèle d'assurance de la qualité](#) ») ne sont pas prises en compte, car elles servent uniquement à montrer à l'IB que les examinateurs ont compris la norme de notation.

Les chefs d'équipe sont des examinateurs très expérimentés à qui l'IB demande d'aider d'autres examinateurs à comprendre la norme de notation. Les chefs d'équipe apportent leur soutien aux examinateurs tout au long du processus de validation et leur font un retour d'information si les copies de contrôle indiquent qu'ils s'éloignent des normes à suivre. Pour chaque matière, le nombre de chefs d'équipe dépendra du nombre d'examineurs requis. Dans le cas d'une matière présentée par un petit nombre de candidats, l'examineur principal apporte ce soutien à tous les examinateurs.

Les examinateurs superviseurs sont des examinateurs expérimentés à qui il est demandé d'aider les examinateurs principaux dans leurs différentes tâches. Il est fréquent qu'ils soient également chefs d'équipe.

Les responsabilités du personnel de l'IB

Chaque matière sera confiée à un responsable de matière chargé de l'évaluation, qui gèrera tout le cycle d'évaluation et s'assurera que les procédures de contrôle de la qualité pertinentes ont été suivies. Beaucoup de responsables de matières sont également des spécialistes de la matière concernée et, de ce fait, ils sont aptes à apporter leur soutien aux examinateurs en chef et aux examinateurs principaux.

Au sein de l'IB, plusieurs équipes gèrent les demandes de modification des épreuves d'examen, de dispositions spéciales ou de mesures compensatoires, les allégations de mauvaise conduite (y compris le plagiat), le contrôle de la qualité du travail des examinateurs et le calcul des facteurs de révision de notation.

Le chef de la direction de l'évaluation, assisté par le responsable en chef de l'évaluation du programme concerné et le responsable en chef des principes et pratiques de l'évaluation, a pour responsabilité d'étudier les seuils d'attribution des notes finales recommandés par l'examineur en chef pour chaque matière et de les accepter ou de demander à l'examineur en chef de revoir ce qu'il a recommandé.

Enfin, il incombe à l'IB de prendre toutes les décisions qui concernent le cycle d'évaluation. Et s'il fait appel à des spécialistes externes pour l'aider à produire des évaluations justes et de qualité, il conserve la responsabilité finale des décisions prises.

Les rôles se rapportant à la rédaction des épreuves d'examen

Les compétences essentielles attendues d'un examinateur principal sont la cohérence dans l'application de la norme de notation et la capacité d'expliquer aux autres examinateurs en quoi consiste la norme.

- Les compétences essentielles requises pour rédiger un examen de qualité sont très différentes et imposent notamment de faire preuve :
 - de créativité, afin de formuler des questions intéressantes et uniques reflétant néanmoins le programme d'études ;
 - d'une connaissance et d'une compréhension des différentes approches pédagogiques et culturelles, afin de formuler des questions dépourvues de biais ;
 - de compétences linguistiques, afin de formuler des questions claires et sans ambiguïté qui conserveront ces caractéristiques une fois traduites dans les autres langues requises ;
 - de clarté, afin de formuler des questions évaluant précisément ce qui est censé être évalué (pertinence de construit) et pouvant être notées de façon cohérente (fiabilité).

L'IB observe souvent que les examinateurs principaux et les examinateurs superviseurs possèdent ces deux ensembles de compétences. Il met donc cette situation à profit pour que les mêmes personnes rédigent et notent les évaluations, avec l'aide des responsables de matières.

Outre les rôles créatifs, d'autres personnes interviennent dans la production des évaluations finales.

- Les agents à la production technique sont chargés de donner aux questions de l'évaluation leur forme finale, à savoir celle d'un examen, que ce soit sur papier ou sur ordinateur.
- Les traducteurs transposent l'évaluation dans les autres langues dans lesquelles elle sera passée.

- Les réviseurs externes sont des spécialistes de la matière qui répondent aux questions de l'évaluation comme s'ils étaient des candidats, puis font un retour d'information sur sa longueur globale et les éventuelles ambiguïtés ou erreurs.

De plus amples informations sur la manière dont les épreuves d'examen sont rédigées sont fournies dans la section « Préparation des épreuves d'examen : élaboration et qualité ».

La hiérarchie des examinateurs

- L'examineur principal est l'arbitre final de la note à attribuer dans une composante donnée. Tous les autres examinateurs doivent suivre cette norme.
- L'IB part du principe que les personnes ayant aidé l'examineur principal à établir la norme ont une meilleure compréhension de cette norme que les personnes qui en ont eu connaissance avec le matériel de normalisation ou auprès de leur chef d'équipe. Cela signifie que les notes des examinateurs étant plus proches de l'examineur principal dans la hiérarchie prévalent sur celles des personnes plus éloignées.
- L'adjectif « superviseur » fait référence à la proximité avec l'examineur principal dans la hiérarchie et non à la durée de l'expérience dans la correction d'examens ou dans l'enseignement.

Le principe qui sous-tend la notation de l'IB est que l'examineur principal est l'arbitre final de la note correcte à attribuer au travail d'un candidat. Même s'il faut reconnaître que les définitions d'une note « correcte » peuvent être très différentes et tout aussi raisonnables les unes que les autres, il n'est pas juste qu'un candidat obtienne une note différente en fonction de l'examineur qui corrige son travail ; c'est pourquoi l'IB demande à l'examineur principal d'établir la norme et à chaque examinateur de s'y conformer.

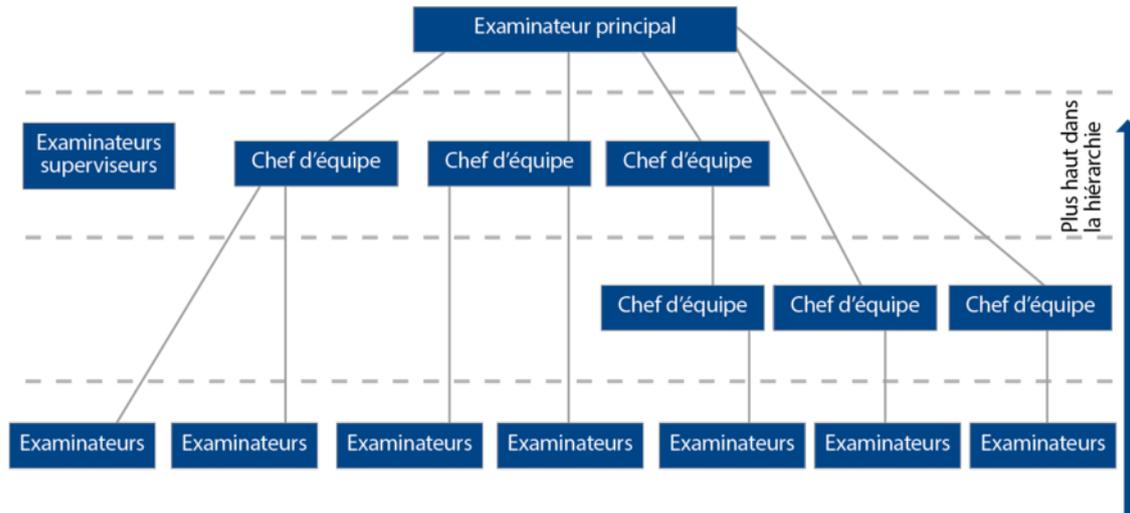
L'examineur principal est aidé par une équipe d'examineurs superviseurs pour établir la norme. Étant donné que ces examinateurs ont l'occasion de discuter de la norme avec l'examineur principal, l'IB considère qu'ils en auront une meilleure compréhension. De ce fait, s'ils contestent la note attribuée par un autre examinateur (ce qui se produit généralement à la suite d'une recorection de copies à risque ou d'une réclamation concernant les résultats), l'IB utilisera la note de l'examineur le plus haut placé dans la hiérarchie comme note définitive du travail en question.

La plupart des chefs d'équipe sont des examinateurs superviseurs, mais, dans le cas contraire, plus de poids est accordé à la note d'un examinateur superviseur qu'à celle d'un chef d'équipe. De même, la note d'un chef d'équipe est préférée à celle d'un examinateur qui a découvert la norme de l'examineur principal dans le matériel de normalisation et dans les instructions de son chef d'équipe.

C'est en ce sens qu'il existe une hiérarchie parmi les examinateurs. Pour une matière comptant un grand nombre de candidats, cette hiérarchie est structurée comme suit.

Figure 39

Hiérarchie des examinateurs pour une matière qui compte un grand nombre de candidats



Cette hiérarchie n'est pas rigide. Si les données issues des copies de validation ou de contrôle montrent qu'un examinateur a une meilleure compréhension de la norme de l'examineur principal qu'un membre placé plus haut dans la hiérarchie, alors cet élément sera pris en considération. Le rôle du chef d'équipe implique d'être capable d'expliquer la norme et pas uniquement de l'appliquer.

Dans l'idéal, nul ne devrait avoir à prendre une décision en se référant aux priorités données par la hiérarchie, car toutes les notations se font selon la même norme. Dans la réalité, cependant, il s'agit d'une méthode fondamentale pour déterminer quelles notes l'IB doit utiliser. Si la situation est particulièrement difficile ou controversée, la copie d'examen est soumise à l'examineur principal, qui décide de la note qui convient.

Intégrité de l'évaluation

- Les évaluations de l'IB ne peuvent être justes que si tous les candidats bénéficient d'une égalité des chances.
- Les différentes formes de mauvaise administration et de mauvaise conduite défavorisent les candidats qui ont suivi les règles ; c'est pourquoi l'IB fait tout son possible pour empêcher ces comportements.

Les règles de l'IB, énoncées dans le *Règlement général* et dans d'autres documents spécifiques à chaque programme, ont été établies pour limiter les possibilités de mauvaise conduite et de mauvaise administration. Toutefois, seul l'établissement scolaire peut créer une culture de l'apprentissage refusant et dénonçant la mauvaise conduite.

- Même si certaines formes d'évaluation se prêtent moins que d'autres à la mauvaise conduite, l'IB garde pour principe que la pertinence de construit (c'est-à-dire le fait d'évaluer ce que l'on souhaite réellement évaluer) doit rester la priorité absolue lors de la conception des évaluations.
- Il est fort probable que l'évaluation sur ordinateur sera d'une grande utilité pour pousser encore plus loin le maintien de l'intégrité des évaluations de l'IB.
- Quiconque a l'impression qu'un cas de mauvaise conduite n'est pas traité par un établissement doit le signaler à l'IB en envoyant un courriel à complaints@ibo.org ou en se mettant en rapport avec le service L'IB vous répond.

L'intégrité intellectuelle est un ensemble de valeurs et de compétences, qui encourage l'intégrité personnelle et les bonnes pratiques en matière d'enseignement, d'apprentissage et d'évaluation. Elle ne doit pas être considérée comme une série de règles strictes à imposer, mais prendre la forme d'une culture au sein de l'établissement scolaire et de la communauté élargie, y compris chez les tuteurs légaux. Bien qu'il soit plus facile d'expliquer aux élèves en quoi consiste le **non-respect** de l'intégrité intellectuelle, en faisant référence à la collusion, au plagiat et à la tricherie, une telle approche ne créera pas la culture positive de l'intégrité qui aboutira naturellement à des résultats d'évaluation plus justes.

Pour être valide, une évaluation doit donner une idée précise de l'accomplissement d'un candidat par rapport à tous les autres candidats qui ont réalisé cette évaluation. C'est pourquoi l'IB met tout en œuvre pour avoir des approches cohérentes de la notation, de l'attribution des notes finales et de l'élimination des biais de ses examens. Les règles et règlements établis à cet effet sont un autre aspect de la création de cette « égalité des chances ».

Le *Règlement général* spécifique à chaque programme définit la mauvaise conduite comme un comportement procurant ou susceptible de procurer un avantage déloyal au candidat qui l'adopte ou à tout autre candidat. Une telle activité affecte non seulement le candidat qui l'entreprend, mais encore toutes les personnes qui passent l'évaluation en même temps que lui, car sa validité s'en trouve réduite. L'IB prend donc la mauvaise conduite très au sérieux, et les mesures de prévention comme ses conséquences sont expliquées en détail dans le document *Procédures d'évaluation* spécifique à chaque programme, ainsi que dans les publications sur l'intégrité intellectuelle.

La conception de l'évaluation joue un rôle capital pour empêcher la mauvaise conduite, car certains types d'évaluations sont plus faciles à contrôler que d'autres. Par exemple, il est plus difficile de demander de l'aide à quelqu'un pour un examen écrit que pour un travail d'évaluation interne. Même si ces éléments sont pris en considération pour définir les approches de l'évaluation, l'IB estime que la pertinence de construit (c'est-à-dire le fait d'évaluer ce que l'on souhaite réellement évaluer) ne doit pas être sacrifiée pour empêcher toutes les possibilités de mauvaise conduite.

La capacité de gestion est un autre aspect essentiel de la validité, et l'IB doit aussi en tenir compte lorsqu'il s'efforce d'empêcher les possibilités de mauvaise conduite. La préparation des salles d'examen, notamment pour l'évaluation sur ordinateur, peut présenter un vrai défi pour les établissements scolaires, et cela ne doit pas être négligé. L'IB s'efforce de penser aux établissements lorsqu'il fixe ses différentes règles, et il tient beaucoup à ce que les coordonnateurs et les chefs d'établissement lui fassent part des pratiques qu'ils ont jugées adaptées ou au contraire compliquées.

Ces échanges permettent également de mieux comprendre les nouveaux problèmes dans le domaine de l'intégrité intellectuelle, et il est donc conseillé aux coordonnateurs et aux chefs d'établissement de transmettre leurs préoccupations ou réflexions par courriel à complaints@ibo.org ou au service L'IB vous répond.

La gestion des conflits d'intérêts

Au sein de l'IB, l'accès aux épreuves d'examen est rigoureusement contrôlé, et l'organisation gère activement tous les liens avec les candidats passant les examens qui pourraient créer un conflit d'intérêts. Si un tel lien est avéré, alors les responsabilités des membres du personnel sont réorganisées.

L'IB part en effet du principe qu'aucun examinateur ne peut noter le travail de ses propres candidats. Dans l'éventualité extrêmement rare où cela est inévitable pour des raisons qui échappent au contrôle de l'IB (par exemple, l'impossibilité de trouver un autre examinateur qualifié pour noter selon la norme requise), un second examinateur indépendant révisé la notation afin de s'assurer que rien n'indique qu'une norme différente a été appliquée aux candidats de l'examineur.

La gestion de la mauvaise administration

Pour chaque programme de l'IB, le *Règlement général* et le document *Procédures d'évaluation* rassemblent les règles et les directives pour minimiser les risques de mauvaise administration.

Pour empêcher tout accès non autorisé, les épreuves d'examen doivent être conservées dans un coffre-fort situé dans une salle fermée à clé. Les épreuves sont expédiées dans des pochettes scellées afin que toute ouverture soit immédiatement visible. La violation des épreuves d'examen est l'une des situations les plus compliquées à contenir pour l'IB, car il est difficile de déterminer la portée de la diffusion des épreuves. Par conséquent, toutes les mesures qui seront prises toucheront généralement un grand nombre de candidats, afin de veiller à ce que l'expérience soit la même pour tous.

Pour que les candidats soient sûrs que les pochettes contenant les épreuves d'examen ont été conservées en lieu sûr avant l'examen, celles-ci doivent être ouvertes devant eux. Si cela n'est pas le cas, les candidats doivent en avvertir l'IB.

Les surveillants doivent rester vigilants pendant toute la durée de l'examen pour s'assurer qu'aucune mauvaise conduite n'est adoptée.

Pour s'assurer de l'observation de ces règles, l'IB procède à des inspections dans les établissements scolaires durant les sessions d'examens. Il ne peut cependant s'agir que de vérifications ponctuelles des procédures. Pour l'IB, il incombe donc essentiellement aux chefs d'établissement et aux coordonnateurs des programmes d'empêcher la mauvaise administration, car ils sont les mieux placés pour garantir que des normes de qualité sont maintenues dans leur établissement jour après jour. Il leur revient en outre de s'assurer que la culture de leur établissement encourage les enseignants et les élèves à adopter les meilleures pratiques qui soient et l'intégrité la plus rigoureuse possible.

Les difficultés posées par les examens internationaux et les zones horaires

L'IB compte des établissements à travers le monde, ce qui le place face à des défis particuliers. Alors que la plupart des systèmes nationaux permettent de faire passer un examen à tous les candidats au même moment, l'IB ne peut adopter une telle pratique sans désavantager certains candidats, en raison des

différents fuseaux horaires. Imaginez par exemple qu'un examen commence à 5 heures du matin ou qu'il finisse après minuit.

La réalité est donc telle que certains candidats finissent leurs examens avant que d'autres les aient commencés. De ce fait, il est demandé aux candidats et aux enseignants de faire preuve d'une intégrité absolue pour ne pas profiter de la situation. Lorsque le grand nombre de candidats justifie une telle mesure, deux épreuves différentes sont prévues en fonction des zones horaires, et ce, afin de réduire le temps entre la fin d'une épreuve pour les candidats d'une zone horaire et le début de cette même épreuve pour les candidats de l'autre zone horaire. Cette solution ne résout cependant pas tout, car il peut exister de grandes différences entre les heures où les épreuves débutent.

Le problème des zones horaires peut être atténué de plusieurs manières. Premièrement, un calcul égoïste peut être fait : pourquoi les candidats choisiraient-ils d'avantager d'autres candidats ? Deuxièmement, il est possible de procéder à un contrôle minutieux des sites Web sur lesquels les épreuves pourraient être diffusées. La règle qui interdit aux candidats de faire sortir les épreuves d'examen de la salle vise à minimiser les risques de partage des questions en ligne.

Troisièmement, l'IB veille à ce que les questions soient formulées de façon à évaluer la compréhension plutôt que le rappel des connaissances. Compte tenu des délais limités, il est en effet peu avantageux de disposer d'un peu plus de temps pour réfléchir sur la réponse lorsque le test ne porte pas sur la vitesse de réflexion (comme c'est le cas avec le rappel des connaissances). Enfin, le fait d'interdire aux candidats de discuter du contenu d'une épreuve d'examen pendant les 24 heures qui suivent un examen supprime toute possibilité de « discussion innocente ».

L'IB cherche constamment à faire preuve d'innovation pour réduire les risques de tricherie dus aux différents fuseaux horaires.

Le plagiat dans les travaux réalisés dans le cadre des cours

Un domaine pose particulièrement difficulté : s'assurer que le travail entrepris en classe (et à la maison) est bel et bien celui du candidat.

L'IB sait parfaitement qu'un certain nombre de sites Web se proposent « d'aider » les candidats pour leur travail. La meilleure défense contre ce type de mauvaise conduite consiste à compter sur l'enseignant qui aura travaillé auprès du candidat, car il pourra déterminer à quels moments un travail n'est pas le reflet du niveau habituel du candidat. C'est pourquoi il est demandé au candidat et à l'enseignant de confirmer que le travail envoyé à l'IB est bien celui du candidat. Le but n'est pas de créer inutilement une complication supplémentaire : l'IB n'est nullement obligé d'accepter une autre version une fois qu'un travail lui est parvenu.

Toutes les écoles du monde de l'IB sont tenues de mettre en place et d'entretenir une culture de l'intégrité intellectuelle, et des manquements répétés entraîneront le retrait de l'autorisation de proposer le programme de l'IB. Il ne suffit pas d'utiliser le logiciel Turnitin : les enseignants doivent en parallèle travailler auprès de leurs candidats pendant qu'ils rédigent leurs travaux d'évaluation interne, afin de vérifier l'authenticité de ces travaux.

Pour ne laisser aucune place au doute, aucun niveau de plagiat n'est accepté, et toutes les citations d'auteurs doivent être correctement référencées, dans le respect des explications données dans les règlements de l'IB. Pour détecter le plagiat, l'IB utilise un certain nombre de logiciels, parmi lesquels Turnitin.

Les avantages de l'évaluation sur ordinateur

L'évaluation sur ordinateur présente plusieurs avantages pour gérer la mauvaise conduite. Ce système réduit en effet les possibilités d'accéder aux épreuves avant les examens, car il se sert du cryptage et de mots de passe. Il permet également de contrôler quand les fichiers d'examen ont été ouverts, et par qui.

Du point de vue de la mauvaise administration, l'IB peut exiger des établissements scolaires qu'ils enregistrent et justifient toutes les modifications apportées, comme le temps supplémentaire accordé ou les pauses autorisées pendant l'examen. Cela signifie que l'IB peut déléguer davantage de responsabilités aux établissements scolaires pour apporter des modifications raisonnables, tout en garantissant l'égalité des pratiques.

Le support électronique permet également de mieux comprendre à quel moment les réponses sont données. Il permet, par exemple, de prouver qu'un candidat avait terminé une question avant la survenue d'un incident rapporté. Il permet aussi une vérification à grande échelle de l'existence de réponses similaires, chose qu'il est impossible d'envisager avec des copies manuscrites.

L'IB reconnaît que les évaluations sur ordinateur seront exposées à de nouvelles formes de mauvaise conduite, notamment à des tentatives de piratage, mais de nouveaux outils sont mis au point pour lutter contre ces attaques.

Les ressources

Les enseignants, les établissements scolaires et les élèves sont les mieux placés pour s'élever contre les cas de mauvaise conduite et y mettre un terme, en créant une culture les refusant et en traitant les cas éventuels avec vigilance. Les documents ci-dessous peuvent aider les établissements à créer une culture de l'intégrité intellectuelle. Ils sont disponibles sur le site Web de l'IB, sous [Outils numériques](#).

- *L'intégrité intellectuelle au sein de l'IB*
- *Savoir citer et référencer ses sources*

Si vous avez besoin d'un soutien supplémentaire ou si vous avez des doutes quant au traitement de certains cas de mauvaise conduite par votre établissement, veuillez envoyer un courriel à complaints@ibo.org ou vous adresser au service [L'IB vous répond](#).

L'équité pour tous : répondre aux besoins des candidats

Pour que les évaluations de l'IB soient valides, aucune discrimination ne doit être pratiquée à l'encontre des candidats ayant des besoins particuliers. À cet effet, l'IB étudiera toute demande de modification des épreuves d'examen et d'aménagement de la procédure d'évaluation à des fins d'inclusion, comme il est expliqué dans le document *Procédures d'évaluation* spécifique à chaque programme.

- Le meilleur moyen de garantir l'équité est de concevoir une évaluation accessible à tous, qui supprime de ce fait tout besoin de modification. Cette idée est englobée dans la notion de conception universelle de l'évaluation, qui fait partie intégrante de l'engagement de l'IB envers la conception universelle de l'apprentissage (CUA).
- Lorsque les besoins des candidats sont connus en avance, ils peuvent donner lieu à des aménagements de la procédure d'évaluation à des fins d'accès et d'inclusion, parmi lesquels la modification des épreuves d'examen.
- D'autres circonstances se présentent à courte échéance ou ne peuvent être gérées au moyen des aménagements de la procédure d'évaluation à des fins d'inclusion. Le cas échéant, des processus de mesures compensatoires sont alors engagés.
- Enfin, le but de tous ces aménagements est de créer une équité pour tous les candidats. Par conséquent, lorsque l'IB doit prendre une décision, il doit tenir compte de ce qui est juste pour l'ensemble de la cohorte et non uniquement pour le candidat concerné. L'objectif est de parvenir à une égalité des chances pour tous les candidats.

L'IB estime que tous les candidats doivent pouvoir passer leurs examens dans des conditions aussi équitables que possible. Il reconnaît que les conditions normales d'évaluation peuvent désavantager les candidats ayant des besoins en matière de soutien à l'apprentissage en les empêchant de montrer le niveau qu'ils ont atteint. De même, il admet que, parfois, des événements ou des circonstances qui échappent au contrôle des candidats auront une incidence sur leurs résultats et donc qu'ils doivent être pris en considération.

Le meilleur moyen de garantir l'équité d'une évaluation est de faire en sorte que tout le monde passe la même évaluation de la même manière. Nombre des modifications apportées pour répondre à des besoins spécifiques pourraient aider tous les candidats à comprendre les questions et à les traiter. Dans l'idéal, l'IB devrait concevoir toutes ses évaluations sur la base d'une compréhension des différents besoins des candidats au lieu de traiter certains candidats différemment. Cette idée est celle de la conception universelle de l'évaluation. L'IB a conscience que cette approche prônant une inclusion totale ne peut pas toujours être adoptée ; c'est pourquoi il est également possible de demander des aménagements de la procédure d'évaluation à des fins d'inclusion.

Les **aménagements à des fins d'accès et d'inclusion** sont prévus pour les candidats qui présentent des besoins individuels tels que :

- des troubles d'apprentissage ;
- des difficultés du langage ;
- des difficultés spécifiques d'apprentissage ;
- des difficultés de communication et des troubles de la parole ;
- des troubles du spectre autistique ;
- des troubles sociaux, émotionnels et comportementaux ;
- des handicaps multiples, des troubles physiques, des troubles sensoriels, des problèmes médicaux ou de santé mentale.

Tous les aménagements raisonnables seront considérés pour un candidat présentant des besoins spécifiques. Pour obtenir de plus amples informations, veuillez consulter le document *Procédures d'évaluation* spécifique à votre programme et la publication de l'IB intitulée *Candidats ayant des besoins en matière d'aménagement de la procédure d'évaluation*.

Les **circonstances défavorables** sont celles échappant au contrôle du candidat et risquant d'avoir des effets négatifs sur ses résultats. Ces circonstances sont étudiées par le comité d'attribution des notes finales et, en cas d'issue favorable, les candidats près du seuil d'attribution des notes finales se verront attribuer la note finale supérieure.

Le principe d'équité pour tous les candidats reconnu par l'IB implique que l'étude des circonstances défavorables et des demandes d'aménagements de la procédure d'évaluation à des fins d'inclusion ne crée pas de situation injuste pour les autres candidats passant les évaluations. L'objectif est de parvenir à une égalité des chances pour tous les candidats.

Principes sous-tendant les aménagements à des fins d'accès et d'inclusion

Les principes qui sous-tendent les aménagements de la procédure à des fins d'accès et d'inclusion sont exposés dans le document de l'IB intitulé *Candidats ayant des besoins en matière d'aménagement de la procédure d'évaluation*. Le texte ci-dessous est extrait de la version consacrée au Programme du diplôme, mais des principes similaires s'appliquent aux autres programmes.

1. L'IB doit garantir qu'une note finale attribuée dans une matière reflète bien le niveau d'accomplissement d'un candidat. Par conséquent, les mêmes normes d'évaluation s'appliquent à tous les candidats, qu'ils aient ou non des besoins en matière de soutien à l'apprentissage.
2. Les aménagements à des fins d'accès et d'inclusion, y compris les ajustements raisonnables, sont des mesures prises en amont des examens pour que l'évaluation soit à la portée du candidat concerné. Ils ne peuvent pas être demandés rétrospectivement, que ce soit pour les épreuves orales ou écrites.
3. Les aménagements demandés pour un candidat ne doivent cependant pas lui offrir un avantage dans une composante d'évaluation.
4. Les aménagements à des fins d'accès et d'inclusion décrits dans le présent document sont destinés aux candidats qui possèdent l'aptitude à satisfaire à toutes les modalités d'évaluation permettant l'octroi du diplôme ou des résultats de cours du Programme du diplôme.
5. Si des aménagements de la procédure à des fins d'accès et d'inclusion sont nécessaires à un candidat pendant ses études dans le cadre du Programme du diplôme ou lors de ses examens d'entraînement, l'établissement est à même de les mettre en place. En revanche, si des aménagements sont requis aux fins de l'évaluation, le présent document dresse la liste de ce qui peut se faire sans l'autorisation préalable de l'IB. Dans toutes les autres situations, il est impératif d'obtenir l'autorisation préalable du centre mondial de l'IB à Cardiff. De la même manière, lorsqu'un candidat du Programme du diplôme a des difficultés pour répondre aux exigences du programme créativité, activité, service (CAS), le service L'IB vous répond doit être consulté.
6. Il est conseillé aux établissements de prévoir des aménagements à des fins d'accès et d'inclusion pour leurs candidats sur la base des critères énoncés dans la présente politique et des observations des enseignants quant au comportement des candidats pendant les travaux et les examens réalisés en classe.
7. Les aménagements à des fins d'accès et d'inclusion demandés pour un candidat doivent correspondre à sa méthode de travail habituelle dans le cadre de ses études. En ce qui concerne les aménagements ne correspondant pas à la méthode de travail habituelle du candidat et n'ayant été mis en place pour soutenir le candidat qu'au cours des six derniers mois de sa scolarité ou ultérieurement, à savoir juste avant les examens, l'IB autorisera de telles demandes d'aménagements à des fins d'accès et d'inclusion uniquement dans des cas très exceptionnels et particuliers.
8. L'IB souhaite accorder des aménagements à des fins d'accès et d'inclusion qui soient compatibles avec ceux normalement en vigueur pour le candidat concerné. Toutefois, seuls les aménagements qui sont

en accord avec la politique et les pratiques de l'IB seront autorisés. Il ne faut pas partir du principe que l'IB acceptera nécessairement les aménagements demandés par un établissement. Les coordonnateurs doivent fournir des informations sur les méthodes de travail habituelles du candidat en classe.

9. L'IB est attaché à une philosophie pédagogique fondée sur la sensibilité internationale. Par conséquent, sa politique en matière d'aménagements à des fins d'accès et d'inclusion ne reflète pas l'usage établi d'un pays particulier. Elle résulte de l'analyse des pratiques en vigueur dans différents pays, pour offrir les mêmes chances aux candidats ayant des besoins en matière d'aménagement de la procédure d'évaluation.
10. L'IB s'efforcera d'offrir, dans la mesure du possible, les mêmes aménagements à des candidats présentant le même type de besoins. Compte tenu des différences culturelles durant la reconnaissance des besoins en matière de soutien à l'apprentissage et de la nature des aménagements de la procédure d'évaluation autorisés dans les établissements, il est parfois nécessaire de trouver des compromis pour permettre une comparaison entre les candidats de différents pays.
11. Toute demande d'aménagements à des fins d'accès et d'inclusion sera jugée au cas par cas. Une autorisation antérieure d'aménagements, par l'IB ou par un autre organisme décernant des titres, n'influencera pas la décision relative à l'autorisation des aménagements demandés par le coordonnateur.
12. L'IB traite toutes les informations relatives à un candidat avec confidentialité. Seuls le personnel de l'IB habilité et les membres du comité d'attribution des notes finales y ont accès si nécessaire et sont tenus de respecter cette confidentialité.
13. Lorsqu'un établissement ne respecte pas les conditions définies par l'IB dans l'octroi des aménagements à des fins d'accès et d'inclusion ou lorsqu'il en accorde sans l'aval de l'IB, le candidat risque de ne pas recevoir de note finale de l'IB pour la matière et le niveau concernés.
14. S'il est possible de prouver que le faible niveau d'un candidat dans sa langue d'usage est dû à un besoin en matière de soutien à l'apprentissage identifié, des aménagements à des fins d'accès et d'inclusion peuvent être accordés. (Pour les matières des groupes 3 à 6, tous les candidats ont le droit d'utiliser un dictionnaire bilingue lors des épreuves écrites.)
15. Les établissements ne doivent pas informer les examinateurs des circonstances défavorables ou des problèmes affectant un candidat (tels que l'autisme, des difficultés rédactionnelles, etc.).
16. Pour le travail évalué en interne, les enseignants ne doivent apporter aucun ajustement dans leur notation.
17. La liste des aménagements à des fins d'accès et d'inclusion disponibles est régulièrement mise à jour. L'IB examinera d'autres aménagements suggérés par un coordonnateur, à condition qu'ils puissent être mis à disposition de tous les autres candidats ayant les mêmes besoins.
18. Le document intitulé *Règlement général du Programme du diplôme* stipule qu'un candidat peut se présenter à trois sessions d'examens pour obtenir le diplôme. Un candidat ayant des besoins en matière de soutien à l'apprentissage peut bénéficier de sessions supplémentaires, à la discrétion de l'IB.
19. Lorsque la nature du trouble d'un candidat ou la nature des aménagements à des fins d'accès et d'inclusion autorisés risquent de gêner les autres candidats pendant un examen, le candidat en question doit passer l'épreuve dans une pièce séparée et être surveillé conformément aux réglementations régissant le déroulement des examens du Programme du diplôme.
20. Les épreuves écrites doivent être surveillées conformément aux réglementations régissant le déroulement des examens du Programme du diplôme. La personne surveillant l'examen du candidat ne doit pas être un membre de sa famille ni quelqu'un avec qui il pourrait y avoir un conflit d'intérêts apparent ou perçu comme tel.
21. Le service L'IB vous répond doit être informé dans les meilleurs délais de toute difficulté liée à la nature des aménagements à des fins d'accès et d'inclusion ou de toute difficulté imprévue éprouvée par le candidat pendant les examens.

Les dispenses d'évaluation

Les dispenses ne sont normalement admises pour aucune composante d'évaluation. Cependant, lorsqu'une composante d'évaluation ou une partie d'une composante d'évaluation demande une fonction physiologique qu'un candidat ne possède pas, ce dernier peut en être dispensé. Toutes les possibilités d'ajustement raisonnable doivent avoir été étudiées avant de déposer une demande de dispense pour une composante d'évaluation. Une autorisation de dispense ne sera accordée que pour des motifs suffisants. L'incapacité physique d'un candidat à réaliser les fonctions demandées pour la composante doit être clairement documentée, et de façon exhaustive.

Pour connaître le détail des principes et des processus liés aux dispenses d'évaluation, veuillez vous reporter au document intitulé *Candidats ayant des besoins en matière d'aménagement de la procédure d'évaluation* correspondant au programme qui vous intéresse.

Les possibilités d'aménagements de la procédure d'évaluation à des fins d'inclusion pour l'évaluation sur ordinateur

Le recours à l'évaluation sur ordinateur permet au candidat d'avoir un plus grand contrôle sur la manière dont l'évaluation est présentée. Les ordinateurs permettent en effet de proposer tout un éventail de polices de caractères, de tailles de texte et de couleurs pour répondre aux besoins précis de chacun, et ce type d'ajustement peut facilement être mis à la portée de tous les candidats.

Dans la plupart des demandes récentes, les aménagements de la procédure d'évaluation à des fins d'inclusion visaient l'autorisation d'utiliser un ordinateur, un besoin qui est à l'évidence comblé par l'évaluation électronique, à condition que le logiciel permettant l'inclusion soit compatible avec l'outil à l'écran. L'IB a parfaitement conscience de cette exigence et fait le nécessaire pour s'assurer que toutes les évaluations sur ordinateur respecteront les normes industrielles associées à de tels logiciels permettant l'inclusion.

Les circonstances défavorables

Les circonstances défavorables ou imprévues sont celles échappant au contrôle du candidat et susceptibles d'être préjudiciables à ses résultats. Ce sont, par exemple, une maladie ou blessure temporaire, un stress grave, des circonstances familiales particulièrement éprouvantes, un deuil ou des événements pouvant menacer la santé ou la sécurité du candidat. Les circonstances défavorables peuvent également inclure des événements qui affectent la communauté scolaire dans son ensemble, comme des troubles civils ou une catastrophe naturelle.

Les circonstances défavorables n'incluent pas les insuffisances dues au fait de l'établissement scolaire, notamment les problèmes liés au personnel enseignant. Il incombe à l'établissement de s'assurer que tous les candidats respectent les exigences du programme et de l'évaluation.

Pour connaître le détail de ce qui est inclus dans la catégorie des circonstances défavorables et ce qui en est exclu, veuillez vous reporter au document *Procédures d'évaluation* et au *Règlement général* du programme qui vous intéresse.

Pour déterminer si un candidat affecté par des circonstances défavorables peut bénéficier de mesures compensatoires, le comité d'attribution des notes finales étudiera les preuves fournies par l'établissement. Si les circonstances qui affectent le candidat sont considérées comme « défavorables », et que le candidat est autorisé à bénéficier de mesures compensatoires, les notes totales du candidat pourront être ajustées dans les matières ou les composantes du programme concernées. Si un ou deux points séparent la note du candidat du seuil d'attribution des notes finales supérieur, sa note finale pour les matières concernées sera augmentée.

La conception universelle de l'évaluation

Les paragraphes précédents s'intéressaient aux situations pouvant amener des candidats à demander des modifications ou une aide particulière pour passer les évaluations de l'IB équitablement, ainsi qu'à la manière dont l'IB gérait ces situations. La meilleure solution reste toutefois d'avoir des évaluations ne dressant pas d'obstacle à la participation. La conception universelle de l'évaluation consiste à prendre en considération l'accès, l'inclusion, l'égalité, les sensibilités culturelles, les stéréotypes et les biais dès le début de la conception de l'évaluation. Elle inclut la création des tâches d'évaluation et des questions d'examen, mais elle remonte également à la conception du modèle d'évaluation global, qui détermine la manière de créer des évaluations comparables chaque session. Si des évaluations plus inclusives sont créées dès le départ, non seulement la pertinence de construit est accrue, mais encore les risques de ne pas satisfaire les besoins de tous les candidats sont minimisés.

La conception universelle de l'évaluation est un aspect de la notion plus globale de conception universelle de l'apprentissage (CUA). La CUA vise à créer des environnements d'apprentissage accessibles à tous les apprenants, notamment aux candidats ayant une déficience, aux candidats de langues ou de cultures différentes, et aux candidats doués et talentueux. Les principes de la CUA peuvent être appliqués à tous les domaines de l'éducation, notamment la conception du programme d'études, la gestion de l'établissement scolaire et l'enseignement. Pour en savoir plus sur la CUA dans le contexte de l'IB, vous pouvez consulter l'ouvrage suivant (en anglais) : RAO, K., CURRIE-RUBIN, R., et LOGLI, C. 2016. *UDL and Inclusive Practices in IB Schools Worldwide*.

Le cycle d'évaluation

- Le processus de création des évaluations doit être imaginé comme un cycle continu dans lequel chaque étape dépend de la précédente et influence la suivante.
- Il faut en moyenne 18 mois pour créer une épreuve d'examen (et son barème de notation), l'IB crée donc des épreuves d'examen pour différentes sessions en parallèle.

L'évaluation s'inscrit dans un modèle cyclique qui compte plusieurs étapes : création, organisation de sessions pour les candidats, notation réalisée par des examinateurs et publication des résultats. La clé de voûte de ce processus est que des enseignements sont tirés de chaque session pour améliorer la qualité de l'évaluation pour la suivante.

Le schéma ci-après n'est qu'une illustration possible de ce processus. Même si de nombreuses étapes pourraient être divisées ou regroupées différemment, il explique de façon adaptée le cycle de vie d'une évaluation. Un résumé de chaque étape et un lien vers la section correspondante sont donnés dans le tableau qui suit le schéma.

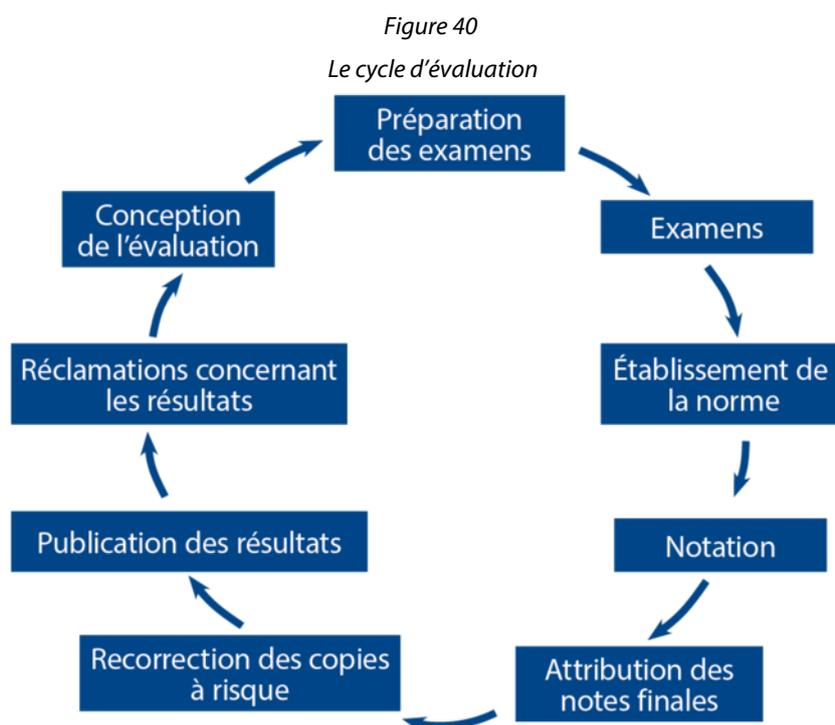


Figure 41
Le cycle d'évaluation

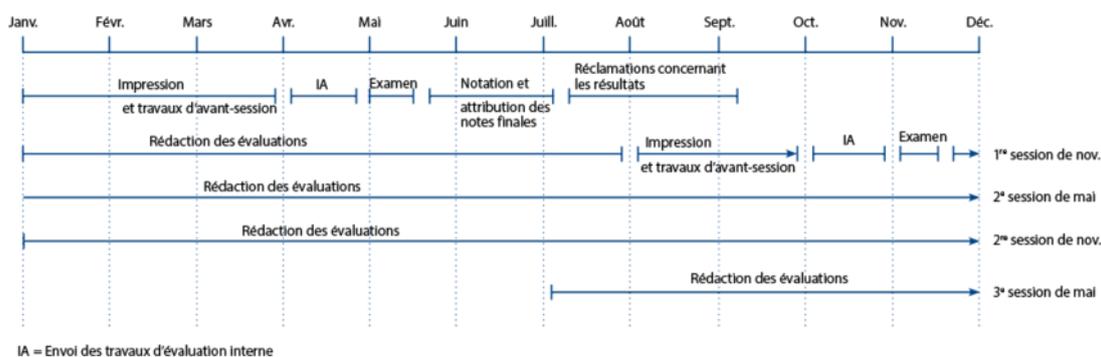
Étape du cycle d'évaluation	Description
Préparation des examens	Processus de création de chaque examen, qui couvre un spectre très large : choix des thèmes abordés dans les questions, rédaction et révision des

Étape du cycle d'évaluation	Description
	questions, traduction dans d'autres langues, modifications diverses pour parvenir à la présentation souhaitée et, enfin, contrôles de la qualité.
Examens	Étape au cours de laquelle les candidats passent les examens dans les établissements scolaires.
Établissement de la norme	Processus au cours duquel les examinateurs superviseurs expliquent à leurs équipes d'examineurs comment noter les travaux des candidats et sélectionnent les copies comportant les notes définitives qui seront utilisées pour vérifier la qualité du travail des examinateurs.
Notation	Activité au cours de laquelle les examinateurs étudient le travail de chaque candidat et décident de la note à attribuer, en suivant impérativement les instructions données par l'examineur principal, la qualité de leur correction étant garantie par des contrôles réguliers.
Attribution des notes finales	Étape au cours de laquelle les examinateurs superviseurs décident de la manière de convertir des notes (qui sont propres à l'examen) en notes finales (qui conservent la même signification, quel que soit l'examen concerné).
Recorrection des copies à risque	Dernier contrôle de la qualité de la notation. Il se concentre sur les domaines dans lesquels des preuves montrent l'existence d'un problème.
Publication des résultats	Diffusion des résultats aux établissements scolaires et aux candidats, mais aussi ensemble de procédures essentiellement administratives : détermination de la réussite du candidat au programme en fonction des résultats obtenus et envoi des relevés de notes aux universités.
Réclamations concernant les résultats	Possibilité offerte aux établissements scolaires d'indiquer à l'IB toute erreur suspectée dans le processus de notation des examens et de demander une vérification du travail du candidat concerné.
Conception de l'évaluation	Aspect le plus important pour clore le cycle d'évaluation. Il consiste à tirer des enseignements des expériences des candidats passant les examens et à améliorer ce que l'on cherche à évaluer et l'approche adoptée, notamment le nombre et le type de tâches d'évaluation.

Pour avoir une idée de la durée de l'élaboration d'une épreuve, veuillez vous reporter au diagramme ci-après, qui indique le temps consacré à chaque étape du cycle d'évaluation. Le cercle représente 3 années complètes et montre qu'il faut en moyenne 18 mois pour préparer une épreuve d'examen. Il montre également que l'IB élabore simultanément les épreuves de mai et de novembre, et celles de l'année suivante.

Figure 42

Le cycle d'évaluation



L'impact de l'évaluation électronique sur le cycle d'évaluation

Le principe du cycle d'évaluation n'est nullement altéré par l'introduction de l'évaluation électronique. Cela permet au contraire de gagner du temps dans certaines étapes du cycle (par exemple, fin de l'expédition par voie postale des épreuves d'examen et des copies d'examen aux centres de numérisation). Il reste toutefois nécessaire de respecter les différentes étapes qui composent le cycle d'évaluation.

Préparation des épreuves d'examen : élaboration et qualité

- Cette étape conduit à des évaluations abouties, prêtes à être passées par les candidats.
- Au cours de sa préparation, une épreuve d'examen doit passer par plusieurs étapes de finalisation :
 - contenu (rédaction de l'épreuve) ;
 - présentation ;
 - facilité d'utilisation ;
 - traduction ;
 - transmission de l'évaluation aux établissements scolaires.
- Toutes les modifications requises par les candidats (aménagements de la procédure d'évaluation à des fins d'accès et d'inclusion) sont également examinées lors de cette étape.

Les règles qui régissent la rédaction des épreuves d'examen

Pour rédiger les épreuves d'examen, l'IB s'appuie sur les qualités d'une bonne évaluation, qui sont décrites dans la section consacrée aux [principes de l'évaluation](#). Pour ce travail, il convient aussi de ne pas perdre de vue l'objectif associé aux évaluations. Les six règles générales suivantes peuvent être considérées comme les éléments essentiels que les auteurs doivent constamment avoir à l'esprit.

- L'authenticité et la pertinence de construit sont les aspects les plus importants de la validité.
- L'évaluation doit donner aux candidats la possibilité de montrer ce qu'ils peuvent faire, et non indiquer ce qu'ils ne peuvent pas faire.
- L'évaluation doit être à la portée du plus grand nombre de candidats possible, tout en permettant une différenciation raisonnable.
- Les meilleures évaluations sont accessibles à tous ; le but est de pouvoir se passer de tout aménagement de la procédure d'évaluation à des fins d'inclusion ([conception universelle de l'évaluation](#)).
- L'évaluation doit porter uniquement sur le programme d'études, tel qu'il apparaît dans le guide pédagogique.
- Les modalités de la notation doivent être décidées lors de l'élaboration de l'évaluation, afin que des points soient bien attribués à ce qui est censé être évalué.

Pour être plus explicite, le barème de notation doit être élaboré en même temps que l'épreuve d'examen, et il faut inclure des orientations sur les marges de tolérance prévues pour les examinateurs.

Les mots-consignes

Les mots-consignes sont les termes et les expressions clés utilisés dans le programme et les questions d'examen pour indiquer ce qui est demandé en réponse à une instruction particulière. Ils suggèrent également le type et le degré de réponse attendus.

Même si les mots-consignes conservent le même sens dans toutes les matières, il est important de s'assurer lors de la révision des épreuves qu'ils sont utilisés conformément à la description qui en est faite dans les guides pédagogiques de l'IB, car les élèves doivent développer une compréhension de ces termes clés dans le contexte précis des matières qu'ils étudient.

Les mots-consignes constituent un défi particulier lorsqu'ils doivent être traduits dans d'autres langues, en raison des subtilités qui peuvent facilement être perdues. Pour que cela ne soit pas le cas, l'IB publie ses mots-consignes dans chaque langue d'usage et explique leur sens. Même si cela peut parfois amener à un emploi inhabituel de certains mots sur le plan linguistique, cela pose moins de problèmes qu'une ambiguïté sémantique dans les évaluations. Lors de la vérification des épreuves traduites, il est donc fondamental de veiller à ce que le bon mot-consigne ait été utilisé.

Les mots-consignes sont largement inspirés de taxonomies reconnues telles que celle de Bloom *et al.* (1956) concernant les objectifs pédagogiques.

Les listes de questions

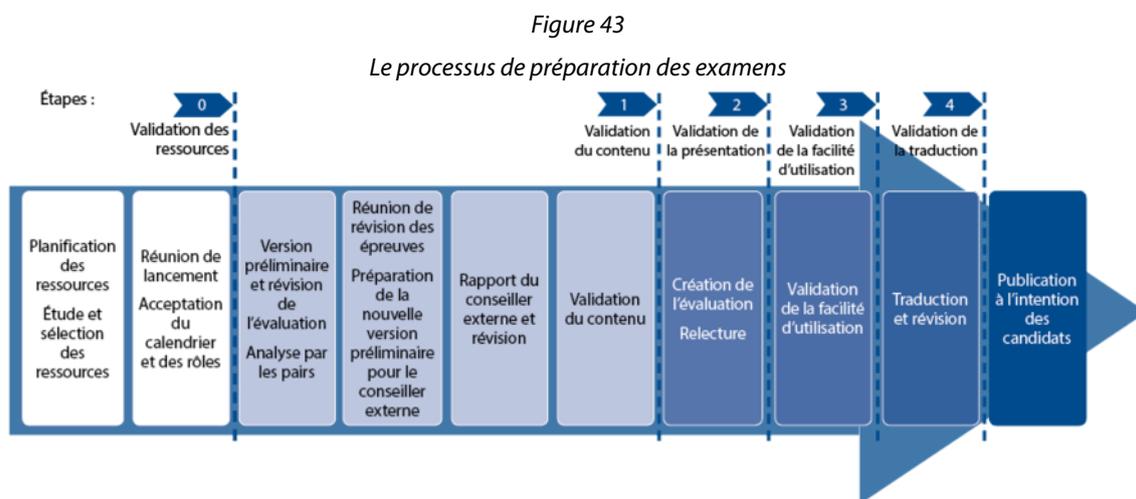
Pour l'heure, l'IB ne se sert d'aucune liste de questions. Chaque épreuve est élaborée comme une évaluation à part entière par ses auteurs.

Dans certains autres systèmes d'évaluation, les questions sont largement mises à l'essai pour déterminer leur exigence cognitive, leur efficacité en tant que discriminateurs, etc. Avec une liste de questions, ces informations sont conservées, ainsi que des détails plus descriptifs tels que le lien entre la question et le programme d'études et le nombre de points, etc.

L'IB ne procède pas à de telles mises à l'essai pour ses évaluations, par crainte d'une fuite dans le domaine public, mais aussi parce que la nature de la plupart de ses questions implique qu'elles ne puissent être jugées convenablement sans une cohorte d'élèves extrêmement conséquente.

Il est probable que l'IB se dirigera à l'avenir vers son propre modèle de listes de questions, peut-être après s'être tourné vers le réseau de collaborateurs de l'IB (IBEN) pour le mettre à contribution. Cela se fera dans le respect du principe de valorisation de l'authenticité et de la pertinence de construit des évaluations.

La présentation de la préparation des épreuves d'examen



Le diagramme ci-dessus illustre les différentes étapes qui constituent le processus de préparation des examens. La création d'une évaluation est un processus lent qui nécessite environ 18 mois au cours desquels interviennent de nombreux spécialistes externes et des membres du personnel de l'IB, dont des auteurs d'épreuves, des spécialistes de l'évaluation, des membres du service d'édition, des vérificateurs externes et des membres du service de traduction.

Il est crucial que le barème de notation soit élaboré en même temps que l'examen. Cette simultanéité garantit que les questions et la notation seront alignées dans ce qu'elles cherchent à évaluer.

Le processus conduisant à la validation des ressources

La première étape de ce processus consiste à déterminer qui rédigera les épreuves. Il sera généralement demandé aux examinateurs principaux de s'en charger, mais après avoir vérifié certains éléments, comme les éventuels conflits d'intérêts en raison de liens avec certains établissements scolaires ou de l'animation d'ateliers de formation. En général, il y a un auteur pour chaque épreuve, et chaque épreuve est ensuite passée en revue par les autres auteurs des épreuves de la matière concernée. En cas de cours de langue présenté par un très petit nombre de candidats, il peut être nécessaire de faire appel à un seul auteur pour toutes les épreuves.

Cette étape aboutit à une « réunion de lancement » au cours de laquelle toutes les personnes prenant part au processus acceptent la responsabilité associée à chaque tâche et les échéances. Voici la liste des différents intervenants.

- Examineur en chef – il peut aussi être auteur d'épreuve (membre externe)
- Auteur de chaque épreuve de la matière – 3 à 5 épreuves par matière en général pour le Programme du diplôme (membre externe)
- Conseiller externe (membre externe)
- Vérificateur (membre externe)
- Éditeur (membre du personnel de l'IB)
- Responsable de matière (membre du personnel de l'IB)

Lorsque l'IB produit des épreuves pour deux zones horaires, deux processus de production des épreuves totalement indépendants sont suivis.

Le processus conduisant à la validation du contenu

À la suite de la réunion de lancement, chaque auteur doit produire une version préliminaire de l'épreuve et une version préliminaire du barème de notation, qui seront ensuite revues par les auteurs des autres épreuves de la matière. Il s'ensuit une réunion officielle de révision des épreuves, au cours de laquelle les éléments suivants sont passés en revue :

- niveau d'exigence cognitive des questions et difficulté globale de l'épreuve ;
- présence éventuelle de biais ;
- respect dans l'épreuve des meilleures pratiques en matière d'accessibilité pour tous ;
- capacité des questions à couvrir le programme d'études ;
- correspondance entre l'épreuve et le modèle d'évaluation publié ;
- lien entre la nouvelle épreuve et les épreuves et échantillons antérieurs ;
- utilisation de questions différentes des exemples donnés dans les ressources publiées.

Cette réunion aboutira à la production d'une autre série de versions préliminaires, qui seront passées en revue par le conseiller externe.

Le conseiller externe est un autre spécialiste de la matière qui n'est pas encore intervenu dans la rédaction de ces épreuves et barèmes de notation. Il porte un « regard neuf » sur ce qui a été produit et doit commenter le niveau de difficulté des épreuves prises ensemble, leur correspondance avec les épreuves antérieures et leur adéquation avec le programme d'études. Le conseiller externe propose des changements auxquels réfléchissent ensuite l'auteur et le responsable de matière.

La validation du contenu définitif de l'épreuve d'examen revient à l'auteur, et le responsable de matière de l'IB doit confirmer que l'épreuve a été passée en revue dans le respect des normes de qualité établies.

Le barème de notation n'est pas encore validé à ce stade : il doit être passé en revue lors de la normalisation, à la lumière des réponses des candidats.

Le processus conduisant à la validation de la présentation

Une fois le contenu validé, le texte des questions ne peut plus être modifié. L'éditeur de l'IB mettra alors les questions au format requis pour les examens.

Pour les examens traditionnels (sur papier), cela signifie ajouter la page de couverture avec les instructions ou les consignes, appliquer les règles stylistiques appropriées, et ajouter les différents codes-barres et pages vierges pour constituer le livret de questions. Ce processus implique également de redessiner certains diagrammes pour obtenir le niveau de qualité d'impression souhaité.

Pour l'évaluation sur ordinateur, l'éditeur aura à créer les examens directement dans l'environnement de développement, qui porte le nom d'IBeADS (acronyme anglais désignant le système d'élaboration des évaluations électroniques de l'IB). Cela vaut également pour les instructions ou les consignes et les styles.

Dans les deux cas, les évaluations sont comparées, après édition, au contenu validé.

Le processus conduisant à la validation de la facilité d'utilisation

Le vérificateur est un autre spécialiste de la matière qui n'est pas encore intervenu dans le processus de production de l'épreuve à ce stade. Il passe l'examen comme s'il était un candidat, en tenant notamment compte du temps pris pour répondre aux questions. Cette vérification a pour but de déceler toutes les erreurs qui ont pu se glisser dans l'examen définitif et de déterminer si les instructions ou les questions comportent des ambiguïtés.

Le vérificateur passe également en revue le barème de notation après avoir terminé l'évaluation.

En cas de cours de langue présenté par un petit nombre de candidats, le rôle du vérificateur est occupé par un second locuteur natif plutôt que par un spécialiste littéraire.

Après examen des commentaires formels du vérificateur, l'évaluation est finalisée. Elle est prête pour l'impression (dans le cas des évaluations sur papier) ou l'inclusion dans le logiciel d'examen (dans le cas des évaluations sur ordinateur) et l'envoi aux établissements scolaires.

Le contrôle de la qualité

Il est essentiel que les épreuves d'examen produites soient exemptes d'erreur et d'ambiguïté, car de tels problèmes pourraient avoir des répercussions considérables sur les candidats qui passent les examens. L'IB prend donc ces questions très au sérieux.

Dans les processus qui ont été décrits précédemment, chaque étape montre qu'il existe un contrôle de la qualité en plus de la validation officielle, qui relève du conseiller externe, du réviseur ou du vérificateur. L'expérience prouve qu'il est préférable de ne pas multiplier les vérifications d'une même erreur. En effet, les contrôles de la qualité sont généralement faits moins scrupuleusement si l'on sait que l'épreuve sera revérifiée par quelqu'un d'autre.

Pour la production des épreuves, l'IB adhère au principe suivant : « Réussir du premier coup et vérifier une seule fois. »

Les traductions

Les candidats doivent pouvoir passer les examens dans la langue utilisée pour l'enseignement. Pour les examens autres que ceux de langue, l'IB leur offre la possibilité de faire leur choix parmi plusieurs langues d'usage. À l'heure actuelle, la plupart des matières sont disponibles en anglais, en français et en espagnol. Des accords ont également été passés localement pour que certaines matières soient proposées en allemand, en japonais ou dans d'autres langues.

Traduire les évaluations n'est pas chose facile. Il est fondamental que le processus de traduction n'altère en rien le sens des questions pour que les candidats ne soient ni avantagés ni désavantagés par le fait de passer l'évaluation dans une langue ou dans une autre. Cela pose des difficultés, car il est nécessaire de reformuler les phrases pour les transposer dans une autre langue. Certains mots ont une signification différente en fonction de la langue, ce qui peut ajouter ou enlever du sens.

La nécessité de traduire les examens est prise en compte dès les premières versions des épreuves, et les équipes d'auteurs sont parfaitement informées des problèmes que cela peut poser. La traduction officielle intervient à la fin du processus, une fois la facilité d'utilisation validée. Les traducteurs engagés par l'IB sont des spécialistes de la matière afin de garantir que le langage technique utilisé dans l'évaluation reste

correct. La traduction est ensuite comparée à l'examen original (et pas uniquement le texte) par un évaluateur bilingue. Il s'agit du dernier contrôle de la qualité.

La gestion des demandes de modification des épreuves d'examen

L'idéal est de concevoir des évaluations adaptées à tous les candidats. Certaines demandes et exigences ne peuvent cependant pas être satisfaites par la simple amélioration de la conception de l'épreuve. Certaines modifications (par exemple, police de caractères ou couleur d'arrière-plan différente et version agrandie de l'épreuve) sont relativement faciles à effectuer. Elles sont apportées lors de la phase de publication des épreuves pour les candidats.

D'autres modifications, comme la transcription en braille, demandent en revanche plus de préparation. Elles peuvent également imposer de mettre au point des tâches connexes, mais différentes, pour pouvoir recueillir des preuves des compétences du candidat dans un domaine. Un exemple typique est une question demandant au candidat de commenter une image, chose qui n'est absolument pas appropriée pour un candidat aveugle. Il convient alors de se demander si une description de l'image est une solution convenable, si cela augmente la charge cognitive du candidat (excès d'informations) ou si cela donne trop d'indications (la description se concentre sur les aspects pour lesquels des points seront accordés). Sinon, est-il possible de formuler la question différemment ?

Ces décisions doivent être prises par l'auteur de l'épreuve, avec l'aide du responsable de l'inclusion et de l'accès à l'évaluation de l'IB. Si nécessaire, l'IB prend également conseil auprès de spécialistes externes. Même si les décisions sont prises au cas par cas, les résultats et les justifications sont systématiquement consignés pour pouvoir être vérifiés à des fins de contrôle de la qualité, et ce, en vue de garantir une cohérence. Lorsque des modifications très spécifiques sont nécessaires, elles doivent être apportées de façon à minimiser les différences entre ce qui est évalué.

Les décisions sont prises après des discussions approfondies et, dans l'idéal, ces dernières doivent avoir lieu avant la validation du contenu. Cela permet non seulement de garantir que l'auteur a toujours à l'esprit les buts initialement fixés pour les questions, mais encore d'améliorer les questions de départ pour supprimer tout besoin de modification. Il faut donc que les établissements scolaires informent l'IB au plus tôt de toute demande d'aménagement de la procédure d'évaluation à des fins d'inclusion.

L'évolution vers une évaluation sur ordinateur

Créer des évaluations sur ordinateur implique des différences considérables dans la manière de préparer chaque évaluation, mais le processus global présenté dans le diagramme reste le même. Il s'agit en effet du processus déjà suivi pour les évaluations sur ordinateur du PEI.

Le plus grand changement concerne l'élaboration des épreuves. Lorsque les évaluations sur ordinateur seront introduites, les auteurs devront comprendre les possibilités qu'elles offrent, afin de faire une utilisation optimale des nouveaux outils qui seront disponibles. Ce changement majeur de ce qui peut être évalué et de la manière de l'évaluer constitue l'un des aspects les plus enthousiasmants de l'évaluation sur ordinateur, mais l'un des plus compliqués aussi.

Les éditeurs travaillant sur les évaluations sur ordinateur devront faire appel à un éventail de compétences différent et plus large. En effet, au lieu de copier le texte dans le bon modèle d'évaluation et de vérifier les exigences stylistiques, ils auront à élaborer des supports multimédias et à s'assurer que l'évaluation dans son ensemble fonctionne comme l'auteur le souhaitait. Le système d'élaboration des évaluations électroniques de l'IB (IBeADS) actuellement utilisé pour le PEI fournit un cadre pour l'évaluation électronique très proche du modèle traditionnel, mais avec des exigences spécifiques qui doivent être maîtrisées par les éditeurs.

La dernière différence majeure concerne la manière dont les évaluations sont transmises aux établissements scolaires. Pour plus d'informations à ce sujet, veuillez vous reporter au [Guide de la session d'examens du PEI](#).

Examens

- Les examens ont pour but de permettre aux candidats de montrer tout ce dont ils sont capables dans un cadre bien délimité présentant une expérience cohérente à tous les candidats et réduisant les risques de mauvaise conduite.
- Lorsqu'il prépare une session d'examens, l'IB doit trouver un équilibre entre les besoins de tous les candidats et la capacité de gestion des établissements, et procéder à la notation le plus rapidement possible pour combler les attentes des candidats.
- L'IB publie des règles claires concernant le comportement attendu dans les salles d'examen, afin de réduire les possibilités de mauvaise conduite. Ces règles sont régulièrement mises à jour pour suivre les évolutions technologiques et les changements d'environnement subséquents.
- Le calendrier des examens est un compromis entre de nombreuses priorités contradictoires, et il constitue l'option « la moins pire » au niveau international.
- Tous les renseignements relatifs aux événements imprévus, aux aménagements à des fins d'accès et d'inclusion, aux circonstances défavorables et à la modification du calendrier, sont donnés dans le document *Procédures d'évaluation* spécifique à chaque programme.

Les examens sont pour les candidats une période stressante pendant laquelle beaucoup est attendu d'eux. Afin de réduire la pression exercée sur eux, les principes de base suivants doivent être respectés :

- limiter la durée des sessions d'examens ;
- éviter, dans la mesure du possible, les chevauchements avec d'autres examens (par exemple, des tests nationaux).

Il est cependant nécessaire de trouver un juste équilibre entre ces éléments et la nécessité d'avoir une évaluation suffisante pour tirer des conclusions valides sur l'accomplissement des candidats et d'empêcher les recours à la mauvaise conduite.

La préparation et la gestion de la salle d'examen

Les établissements scolaires doivent organiser les examens en respectant strictement le règlement présenté dans le document *Déroulement des examens du Programme du diplôme de l'IB* publié pour chaque session, disponible sur le [Centre de ressources pédagogiques](#) dans la section **Mise en œuvre > Procédures et processus pour l'évaluation**. Ce règlement couvre tous les aspects de l'organisation des examens, de la réception du matériel d'examen à l'envoi des copies à un centre de numérisation, en passant par le déroulement des examens.

L'IB accorde une grande importance à la conservation des épreuves d'examen en lieu sûr et au bon déroulement de chaque examen, car, si les établissements scolaires ou les parties prenantes ont le sentiment qu'il y a eu mauvaise conduite, alors la valeur des résultats des candidats s'en trouvera considérablement diminuée. Afin de maintenir un climat de confiance global, l'IB étudie de près tous les cas de mauvaise administration rapportés, et des inspections aléatoires sont effectuées dans les établissements scolaires par des membres de l'IB ou des consultants pendant la session d'examens, afin de vérifier que les épreuves d'examen sont conservées en lieu sûr et que les examens se déroulent conformément au règlement de l'IB. Il ne peut en réalité s'agir que de vérifications ponctuelles des procédures. Pour l'IB, il incombe donc aux chefs d'établissement et aux coordonnateurs des programmes d'empêcher la mauvaise administration, car ils sont les mieux placés pour garantir que des normes de qualité sont maintenues dans leur établissement jour après jour. Il est également attendu que la culture de l'établissement encourage les

enseignants et les élèves à adopter les meilleures pratiques qui soient et l'intégrité la plus rigoureuse possible.

L'un des pires scénarios possibles pour les candidats est de passer les examens, puis que leur établissement ou les services postaux égarent leur travail. Lorsqu'une telle situation se produit, l'IB tente de limiter les dégâts en estimant une note pour les candidats, ce qui n'est envisageable que s'il existe d'autres preuves (épreuves) sur lesquelles s'appuyer. Et pour s'en assurer, l'IB demande à ce que les différentes réponses (copies d'examen) ne soient pas toutes envoyées au centre de numérisation le même jour, afin de réduire la possibilité de perdre toutes les copies durant leur acheminement.

Ce risque sera encore réduit avec les évaluations sur ordinateur, puisqu'une copie du travail du candidat sera mise à disposition à chaque étape du processus de chargement.

L'établissement du calendrier des examens

Parce que l'IB souhaite que la session soit la plus courte possible, plusieurs matières sont prévues au même moment dans le calendrier des examens. Même si tout est fait pour tenir compte des différentes combinaisons de matières possibles pour les candidats, force est de reconnaître qu'il est inévitable qu'un petit nombre de chevauchements se produisent, et la procédure à suivre dans de tels cas est indiquée dans le document *Procédures d'évaluation* spécifique à chaque programme.

Le calendrier des examens est établi de manière à éviter que les candidats aient plus de six heures et demie d'examens dans la même journée. Généralement, les deux ou trois épreuves d'une matière aux composantes multiples ont lieu les unes à la suite des autres, la première épreuve commençant dans l'après-midi et la dernière se terminant le lendemain matin. Bien qu'elle ne puisse pas toujours être mise en place, cette option est préférable à l'organisation de tous les examens d'une même matière le même jour.

Pour la plupart des candidats, cela permet d'échelonner les examens de manière plus équilibrée sur les trois semaines du calendrier des examens et leur donne l'occasion de se reprendre s'ils ont le sentiment qu'ils ne se sont pas montrés à leur juste valeur lors d'un examen donné.

Les principes de conception du calendrier des examens

Il n'est pas toujours possible de suivre tous ces principes. Le cas échéant, un compromis doit être trouvé. L'IB publie le calendrier des examens au moins un an avant leur tenue.

Les points suivants donnent un aperçu général des principes qui sous-tendent l'établissement du calendrier.

- Il est actuellement impossible de prendre en considération les jours fériés, les fêtes nationales, les vacances scolaires ou les fêtes religieuses, en raison du nombre de pays où sont situés les établissements scolaires qui proposent un programme de l'IB.
- Il serait souhaitable de ne pas organiser d'examen le jeudi et le vendredi, par égard pour les établissements scolaires du Moyen-Orient où ces jours sont instaurés comme jours de repos, mais il est actuellement impossible d'agir autrement, et ce, pour des questions pratiques.
- Tout sera mis en œuvre pour tenir compte des liens particulièrement étroits qui peuvent exister entre certaines matières et une région ou une culture en particulier. Par exemple, les examens de littérature ou de langue arabe ne seront pas prévus un vendredi.
- L'IB se sert des données d'inscription relatives aux combinaisons de matières pour s'assurer que les chevauchements touchent un minimum de candidats à travers le monde.
- Les candidats ne peuvent pas être interrogés le même jour sur deux langues étrangères différentes (cette mesure ne concerne pas les langues d'usage).
- Afin de réduire les risques qu'un événement inattendu désavantage un candidat dans toutes les composantes des matières qui comptent plusieurs épreuves, ces dernières seront prévues sur deux jours ou plus.
- Les candidats doivent pouvoir concentrer leurs révisions dans chaque matière dans un espace de temps limité. De ce fait, les épreuves qui s'étalent sur plus d'une journée seront prévues sur des jours consécutifs, dans la mesure du possible.

- Les épreuves des candidats doivent être réparties sur toute la période d'examens et non être concentrées sur une période plus courte. Cela signifie que, dans la mesure du possible, les matières qui comptent le plus grand nombre de candidats (par exemple, l'histoire, l'anglais ou les mathématiques) ne seront pas prévues sur des jours consécutifs.
- Pour la même raison, l'IB s'efforcera de prévoir les examens de langue et de sciences dans chacune des trois semaines du calendrier des examens.
- Pour pouvoir satisfaire à des exigences de traitement des examens internes à l'IB, certaines matières devront avoir lieu au début du calendrier. C'est notamment le cas des matières qui comptent un grand nombre de candidats.

La gestion des événements imprévus

Le document *Procédures d'évaluation* spécifique à chaque programme explique comment gérer la plupart des événements imprévus. Si vous avez le moindre doute à ce sujet, n'hésitez pas à contacter le service [L'IB vous répond](#). Il saura vous apporter tous les renseignements nécessaires.

Les aménagements à des fins d'accès et d'inclusion et les circonstances défavorables

Ces deux points sont largement traités dans la section « [L'équité pour tous : répondre aux besoins des candidats](#) ».

L'IB estime que tous les candidats doivent pouvoir passer leurs examens dans des conditions aussi équitables que possible. Les conditions normales d'évaluation peuvent désavantager les candidats ayant des besoins en matière de soutien à l'apprentissage en les empêchant de montrer le niveau qu'ils ont atteint. Le cas échéant, des aménagements à des fins d'accès et d'inclusion peuvent être autorisés. Il peut s'agir de dispositions spéciales pour passer les examens ou de modifications apportées aux épreuves.

Les événements échappant au contrôle du candidat, qui se produisent au cours de la session d'examens et qui risquent d'avoir des conséquences négatives sur son travail, portent le nom de « circonstances défavorables ». Ces circonstances incluent les maladies ou blessures temporaires, le stress grave, les circonstances familiales particulièrement éprouvantes, le deuil ou les événements pouvant menacer la santé ou la sécurité du candidat. Les circonstances défavorables peuvent également inclure des événements qui affectent la communauté scolaire dans son ensemble, comme des troubles civils ou une catastrophe naturelle. Elles n'incluent pas les insuffisances dues au fait de l'établissement scolaire. Dans de telles situations, l'établissement scolaire doit contacter l'IB et lui indiquer le détail des circonstances, afin que le comité d'attribution des notes finales puisse déterminer si ces circonstances doivent être prises en considération lors de l'établissement de la note finale du candidat.

Pour obtenir davantage d'informations sur ces aménagements, veuillez vous reporter au document *Procédures d'évaluation* spécifique à chaque programme.

La modification du calendrier des examens

Modifier la date d'un examen présente des risques importants pour l'intégrité du déroulement de l'examen en question. Cela implique en effet que des candidats passeront l'examen bien avant ou après d'autres, une situation qui peut très facilement conduire à une mauvaise conduite. Il ne sera jamais accepté qu'un examen soit avancé à la veille du jour initialement prévu, car le risque d'infraction au règlement pour la majorité des candidats passant l'examen serait bien trop élevé.

L'IB peut autoriser un candidat à passer un ou plusieurs examens à une heure ou à une date différentes de celles indiquées dans le calendrier des examens de l'IB dans trois cas seulement :

- des chevauchements entre des examens de l'IB prévus à la même heure et à la même date ;
- des chevauchements entre le calendrier des examens de l'IB et les examens d'un autre organisme décernant des titres, y compris les examens d'entrée à l'université ;

- des situations d'urgence.

La modification de l'heure ou de la date d'un examen exige que le coordonnateur de l'établissement puisse garantir que les sujets d'examen ne seront pas divulgués. Si l'IB autorise l'établissement à avancer l'examen à une heure antérieure ou à le repousser à une heure ultérieure le même jour, le coordonnateur est tenu de s'assurer que le ou les candidats concernés seront surveillés en permanence entre l'heure initialement prévue pour l'examen et la nouvelle heure. Cela est nécessaire pour rendre impossible toute communication avec un autre candidat ayant déjà passé le même examen.

La procédure à suivre pour demander une modification du calendrier des examens est expliquée en détail dans le document *Procédures d'évaluation* spécifique à chaque programme.

Établissement de la norme : comment préparer les examinateurs à la notation

L'établissement de la norme vise essentiellement :

- pour l'examineur principal, à fixer la norme d'évaluation (après en avoir discuté avec les autres examinateurs) à partir du travail des candidats ;
- à tester et à parfaire le barème de notation pour s'assurer qu'il permet de récompenser équitablement les efforts des candidats ;
- à produire des copies comportant les notes définitives (copies d'entraînement, de validation et de contrôle) ;
- à communiquer une compréhension et à la partager avec tous les examinateurs ;
- à confirmer la compréhension de la norme à travers tout le processus de « validation » des examinateurs.

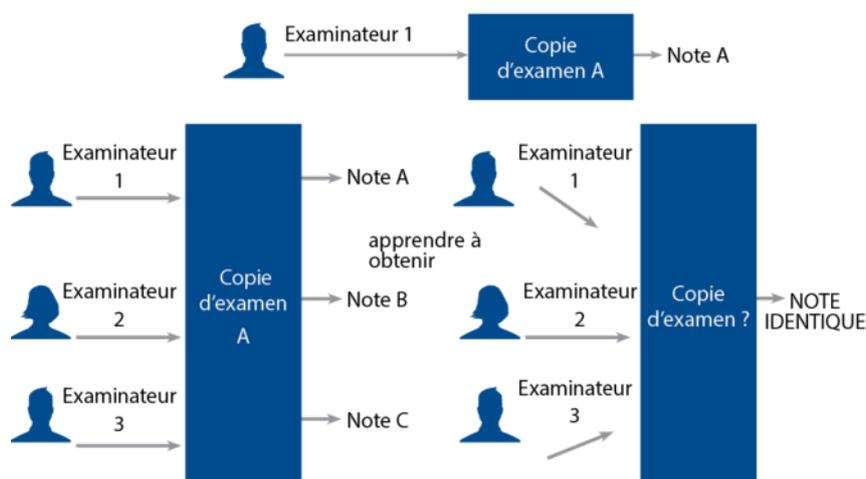
Le terme « établissement de la norme » désigne la période qui couvre plusieurs parties du processus d'évaluation. Cette étape a pour objectif de préparer les examinateurs de façon à ce qu'ils puissent procéder à une notation efficace et fiable.

Il est important que chaque examinateur attribue la même note à des travaux de même qualité, sinon la note finale du candidat sera influencée par l'examineur qui corrigera son travail. Pour chaque composante, une norme est établie par l'examineur principal et tous les autres examinateurs doivent s'y conformer.

Ce processus, qui consiste à établir une même norme pour tous les examinateurs, porte le nom de « normalisation ».

Figure 44

La notation de tous les examinateurs doit être conforme à la norme établie par les examinateurs principaux.



Le processus de normalisation commence par le recrutement d'examineurs possédant l'expérience et les compétences nécessaires. Une formation générale leur est ensuite fournie, afin qu'ils comprennent le processus et ses principes. Vous trouverez de plus amples informations sur le recrutement et la formation des examinateurs sur le [site Web de l'IB](#).

La réunion de normalisation

La discussion au cours de laquelle l'équipe d'examineurs superviseurs établit la norme, sous la direction de l'examineur principal, porte le nom de « réunion de normalisation ». Bien que ces échanges puissent se faire de façon traditionnelle, il est courant de recourir aux forums de discussion de l'IB et à la visioconférence pour éviter aux examinateurs superviseurs, qui sont répartis à travers toute la planète, d'avoir à se rendre à Cardiff. Un tel déplacement ne sera exigé que pour des raisons particulières, comme la première session d'examens d'un nouveau cours ou un processus de normalisation particulièrement compliqué.

Après les examens, les examinateurs superviseurs passent en revue des exemples de travaux de candidats pour s'assurer que le barème de notation couvre convenablement tout l'éventail des réponses que les examinateurs pourront rencontrer. Même si toutes les réponses possibles ne peuvent pas être couvertes, l'examineur principal et l'équipe d'examineurs superviseurs doivent veiller à ce que le barème de notation explique clairement ce qui est attendu pour chaque note.

Il est important que le barème de notation ou les remarques sur la notation ne soient pas détaillés au point de perdre en efficacité, parce que les examinateurs ne suivent plus ou passent à côté de certains éléments fondamentaux. L'inverse est également vrai : un barème de notation doit contenir suffisamment d'informations pour garantir que tous les examinateurs accorderont le même nombre de points à certaines réponses. À titre indicatif, si une remarque particulière est pertinente pour moins de 10 % des candidats, alors il sera certainement inutile de l'inclure dans les orientations.

La réunion de normalisation mène notamment à la constitution d'une série de **copies comportant des notes définitives**. Il s'agit d'exemples de travaux de candidats qui ont été notés par l'examineur principal, et qui seront ensuite utilisés pour former et évaluer d'autres examinateurs lors d'un processus qui porte le nom de « modèle d'assurance de la qualité ». Il est indispensable que les copies comportant les notes définitives soient notées correctement, car elles seront utilisées pour émettre un jugement sur la qualité du travail d'autres examinateurs. Les copies comportant les notes définitives sont utilisées à des fins différentes, d'où un classement selon trois catégories :

1. Les copies d'entraînement
2. Les copies de validation
3. Les copies de contrôle

Les copies d'entraînement

- L'objectif des copies d'entraînement est d'aider les examinateurs dans leur apprentissage de la norme de notation.
- Ces exemples doivent montrer comment noter les travaux de candidats typiques ou toute situation courante dans laquelle les examinateurs comprennent mal le barème de notation.

Les exemples de travaux de candidats comportant les notes définitives sont utilisés par les examinateurs pour vérifier leur compréhension du barème de notation et des remarques sur la notation. Une fois ces copies passées en revue, les examinateurs devraient avoir la certitude qu'ils peuvent procéder à une notation correcte.

Lors de la sélection des copies d'entraînement, les examinateurs principaux doivent tenir compte des éléments suivants :

- l'objectif étant d'aider les examinateurs à se préparer au processus de notation, il est inutile d'inclure des copies d'examen qui illustrent une situation qu'ils ne sont pas susceptibles de rencontrer à nouveau ;
- la nécessité d'expliquer les notes attribuées, à l'aide de commentaires informatifs ;
- l'utilité ou la pertinence d'exclure certaines parties de la copie d'examen d'un candidat ou d'ajouter une remarque indiquant qu'elles ne sont pas utiles à l'apprentissage des examinateurs.

Pour être de qualité, une série de copies d'entraînement doit inclure les éléments suivants (par ordre de priorité) :

1. des exemples montrant comment noter un travail typique ;
2. toutes les situations courantes dans lesquelles les examinateurs comprennent mal le barème de notation ;
3. une large gamme de notes, allant de faibles à élevées, pour aider les examinateurs à reconnaître où ils peuvent attribuer des points et comment distinguer les bonnes (et les mauvaises) réponses ;
4. des exemples d'exceptions et de réponses complexes, et comment les traiter.

Les copies de validation

- L'objectif des copies de validation est de montrer ou de prouver que les examinateurs notent selon la norme de notation adoptée.
- Ces copies comprennent des exemples des réponses les plus courantes données par les candidats sur l'ensemble de l'éventail de notes.

Avant d'être autorisés à commencer à noter des travaux de candidats « en conditions réelles », les examinateurs doivent montrer qu'ils sont capables de noter selon la norme adoptée. Pour ce faire, ils sont « évalués » à l'aide de cinq exemples de travaux de candidats déjà notés par l'examineur principal, ce qui permet de savoir au préalable quelle note ils sont censés attribuer.

Si la note attribuée est différente de celle de l'examineur principal, un retour d'information est fait aux examinateurs pour leur indiquer la note correcte et les raisons de cette attribution. Après avoir réfléchi sur ces éléments, les examinateurs se voient offrir une seconde possibilité de montrer qu'ils comprennent à présent la norme de notation.

Pour être de qualité, une série de copies de validation doit inclure les éléments suivants (par ordre de priorité) :

1. des exemples des réponses les plus courantes données par les candidats sur l'ensemble de l'éventail de notes (élevées, moyennes et faibles) ;
2. des exemples de travaux exigeant de l'examineur qu'il ait compris toutes les exceptions courantes du barème de notation ou les erreurs courantes que les candidats sont susceptibles de faire.

Lors de la sélection des copies de validation, les examinateurs principaux ne doivent pas inclure des copies d'examen destinées à « piéger » les examinateurs. Ils doivent choisir des copies difficiles à noter, qui sont représentatives du type de difficultés auxquelles les examinateurs seront confrontés de façon quotidienne. Ils doivent également chercher des exemples de tous les cas de figure importants que les examinateurs doivent être en mesure de traiter correctement.

Les copies de contrôle

- L'objectif des copies de contrôle est de montrer ou de prouver que les examinateurs continuent de noter selon la norme adoptée.
- Ces copies sont représentatives des erreurs courantes que peuvent faire les examinateurs et contiennent des exemples de réponses sur l'ensemble de l'éventail de notes.

Même si un examinateur ne peut pas être autorisé à commencer à noter des travaux de candidats en conditions réelles avant d'avoir compris la norme de notation, il est évident que sa manière de noter peut évoluer au fil du temps. Par conséquent, sa notation est régulièrement vérifiée avec des copies de contrôle.

Les copies de contrôle ont déjà été notées par l'examineur principal, ce qui permet de savoir au préalable quelles notes sont censées être attribuées. Ces copies ressemblent à tous les autres travaux des candidats. L'examineur ne sait donc pas qu'il corrige une copie de contrôle.

Si l'examineur donne la note de l'examineur principal à une copie de contrôle, il peut poursuivre son travail de notation. Toutefois, s'il montre qu'il ne respecte plus la norme de notation attendue, une intervention est nécessaire.

Cela commence par un retour d'information et des orientations, afin de laisser la possibilité à l'examineur de retrouver la norme de notation. Si tel est le cas, l'examineur peut ensuite reprendre sa notation, sachant que tous les candidats pourront obtenir une note cohérente, quelle que soit la personne qui corrige leur travail.

Si l'examineur est incapable de retrouver la norme de notation attendue, il ne pourra plus corriger de travaux de candidats.

En règle générale, les examinateurs doivent corriger une copie de contrôle après dix copies notées en conditions réelles. Étant donné qu'il ne s'agit pas forcément de la dixième copie, les examinateurs ne savent pas à quel moment ils corrigent une copie de contrôle. Si la situation l'exige, la fréquence des copies de contrôle peut être modifiée.

Pour être de qualité, une série de copies de contrôle doit inclure les éléments suivants (par ordre de priorité) :

1. des travaux de candidats représentatifs des erreurs courantes que peuvent faire les examinateurs ;
2. une large gamme de notes (élevées, moyennes et faibles).

Contrairement aux copies de validation, les copies de contrôle sont réparties de façon aléatoire. Chaque examinateur les recevra donc dans un ordre différent.

Lors de la sélection des copies de contrôle, les examinateurs principaux ont bien conscience qu'à ce stade, les examinateurs corrigent des travaux de candidats en conditions réelles. Par conséquent, l'important est de s'assurer que les notes attribuées aux candidats sont correctes. Il est donc approprié d'inclure des copies complexes ou difficiles à noter. Cela n'est pas effectué dans le but de « piéger » les examinateurs, mais de s'assurer que ces derniers respectent la norme de notation. Il est également possible d'utiliser des copies de contrôle afin de vérifier que les examinateurs suivent toutes les règles de notation, et pas uniquement que leur notation est correcte. Même si cette situation est rare, il peut arriver que l'examineur principal choisisse une copie de contrôle qui doit remonter jusqu'au chef d'équipe pour une raison ou une autre (par exemple, des preuves d'une mauvaise conduite ou une réponse totalement inattendue).

Les chefs d'équipe doivent superviser les examinateurs et leur apporter leur soutien tout au long de la notation. Dès qu'un examinateur est contraint d'arrêter de noter, car une de ses copies de validation ou de contrôle a été notée de façon incorrecte, son chef d'équipe doit se mettre en rapport avec lui pour lui faire un retour d'information et le conseiller. Le chef d'équipe et l'examineur pourront consulter la copie de contrôle qui n'a pas été notée selon la norme, ainsi que la note attribuée par l'examineur principal et ses annotations. Le rôle du chef d'équipe est d'expliquer à l'examineur pourquoi la note attribuée n'était pas correcte. Ils pourront ensuite discuter ensemble du barème de notation et de la bonne façon de l'appliquer. Les examinateurs qui continueront à appliquer le barème de notation de façon incorrecte, malgré la formation complémentaire assurée par le chef d'équipe, seront définitivement déchargés de leur travail de notation.

La marge de tolérance

- La marge de tolérance reflète les différences légitimes qui peuvent exister entre les notes attribuées à un même travail par des examinateurs différents. Imaginez, par exemple, que deux enseignants de votre établissement corrigent le même travail. Ils conviennent tous deux de la bonne qualité du travail, mais le premier attribue 46 points au travail et le second 47 points.
- L'IB s'engage à poser des questions qui évaluent ce qui est important, et pas simplement ce qui est facile à corriger.

Pour les questions appelant une réponse qui est soit bonne soit mauvaise, l'examineur doit attribuer exactement la même note que l'examineur principal. Pour les autres questions, notamment celles qui évaluent la compréhension ou l'analyse, un certain degré de jugement intervient dans la note qui sera

attribuée. Deux examinateurs spécialisés ayant la même conception d'une bonne réponse attribueront des notes proches, qui pourront toutefois varier d'un ou deux points.

L'idée de marge de tolérance est utilisée pour rendre compte de ces différences d'opinion légitimes lors de l'examen du travail de notation des examinateurs. Par exemple, si l'examinateur principal a attribué une note de 46 à une dissertation et qu'une marge de tolérance de 2 points était associée à la question, alors il est considéré que tout examinateur attribuant une note comprise entre 44 et 48 respecte la norme de notation.

Pour les questions en plusieurs parties ou notées à l'aide de plusieurs critères, il convient de surveiller les différences au niveau de la note globale (c'est-à-dire totale) mais aussi au niveau de chaque partie ou critère. Un examinateur doit se trouver dans la marge de tolérance pour ces deux aspects pour montrer qu'il a respecté la norme de notation.

La réussite de l'établissement de la norme

Pour savoir si cette étape du cycle d'évaluation a été réussie, il faut vérifier la présence des éléments suivants :

- cohérence de la notation ;
- capacité des chefs d'équipe à intervenir et à aider les examinateurs pour presque toutes les questions relatives à la notation de la composante concernée ;
- instructions, y compris barème de notation et remarques sur la notation, expliquant clairement les motifs de l'attribution des notes ;
- copies comportant des notes définitives adaptées à l'objectif visé (et correctement notées).

Notation

- Pour être réussie, la notation doit permettre aux candidats d'obtenir un résultat cohérent et exact reflétant la qualité de leur travail.
- Les opinions exprimées par l'examineur principal sont correctes et doivent être reproduites par tous les autres examinateurs dans leur notation.
- Notes et notes finales sont deux choses différentes : les candidats peuvent obtenir une meilleure note à un examen plus facile, mais ils doivent se voir attribuer la même note finale.

Comprendre la notation comme le résultat du jugement cohérent des examinateurs

L'objectif du barème de notation est d'apporter un soutien aux examinateurs et de leur rappeler la norme de notation. Le barème de notation n'indique pas à quel endroit attribuer les points. Le jugement de l'examineur principal définit une note « correcte ». Il s'appuie ensuite sur le barème de notation pour expliquer son raisonnement.

Il s'agit là d'un principe fondamental, car cela signifie que si l'opinion exprimée par l'examineur principal va à l'encontre du barème de notation ou des remarques sur la notation, alors le barème de notation est incorrect et doit être modifié.

Fort heureusement, de tels cas sont très rares. Ils sont cependant extrêmement graves, car ils impliquent qu'il faudra réexpliquer aux autres examinateurs qui se sont fiés au barème de notation quelle devrait être la norme de notation correcte. L'IB aura aussi à recorriger tous les travaux concernés par cette erreur, afin qu'aucun candidat ne soit noté de façon incorrecte. Les examinateurs principaux et leurs équipes d'examineurs superviseurs accordent la plus grande attention à l'établissement de la norme afin de garantir que le barème de notation est correct et informatif, et d'éviter ainsi ce type de situation.

Pour être autorisés à noter des copies d'examen, les examinateurs doivent avoir noté avec succès les copies de validation et de contrôle. En effet, il ne leur suffit pas de savoir interpréter le barème de notation pour obtenir l'autorisation de corriger des travaux de candidats en conditions réelles. De même, les enseignants et les parents ne doivent pas tenir pour acquis que leur interprétation du barème de notation certifie qu'une copie d'examen a été notée de façon incorrecte. Dans de tels cas, ils sont invités à faire appel au service des réclamations concernant les résultats.

Les « notes définitives »

Les termes « note définitive » ou « copie comportant une note définitive » impliquent que l'examineur principal a émis un jugement formel sur la note qui doit être attribuée. Les copies comportant les notes définitives sont ensuite utilisées dans le cadre du modèle d'assurance de la qualité.

Notes et notes finales sont deux choses différentes

Pour bien conduire des évaluations sommatives et en tirer profit, il est important de comprendre la différence entre la notation du travail des candidats et l'attribution d'une note finale.

- Dans la notation, le travail produit par un candidat est récompensé au regard d'un barème de notation ou d'un autre cadre similaire. Il s'agit d'une indication de la proportion de la tâche d'évaluation que le candidat a réussie. La note en tant que telle n'a pas d'autre portée.

- En attribuant une note finale, l'examineur émet un jugement sur la qualité du travail du candidat au regard d'une norme définie et en prenant en compte la difficulté de la tâche ainsi que la proportion de la tâche qui a été réalisée. Par conséquent, la note finale a un sens ou une pertinence et doit en principe pouvoir être comparée aux résultats obtenus dans les autres évaluations.

Figure 10

La notation peut être comparée à la qualité (saveur) de la nourriture préparée, mais la note finale rend quant à elle compte de la complexité de la préparation culinaire.



Un candidat pourrait obtenir une note finale élevée en réussissant uniquement une proportion limitée d'une question très difficile, mais ne pas réussir à obtenir cette même note finale en répondant correctement à de nombreuses questions bien moins complexes.

Comme il est expliqué par la suite, il n'est pas nécessaire de faire des références aux aspects réussis par le candidat pour expliquer le niveau correspondant à la note finale, même si l'IB suit une telle approche. En effet, il existe d'autres systèmes tout à fait cohérents et très réputés, dans lesquels le niveau d'un candidat est déterminé par rapport au résultat de ses pairs.

Dans ses évaluations, l'IB se sert habituellement des notes pour avoir une idée des résultats globaux (système de compensation). Il s'intéresse ensuite à ce qui est exprimé par les candidats ayant obtenu ce nombre de points pour fixer une limite (seuil d'attribution des notes finales) au-delà de laquelle attribuer une certaine note finale aux candidats ayant obtenu plus de points. Ce processus est expliqué plus en détail dans la section « [Les pratiques de l'évaluation de l'IB](#) ».

La notation sur ordinateur (et ses principes fondamentaux)

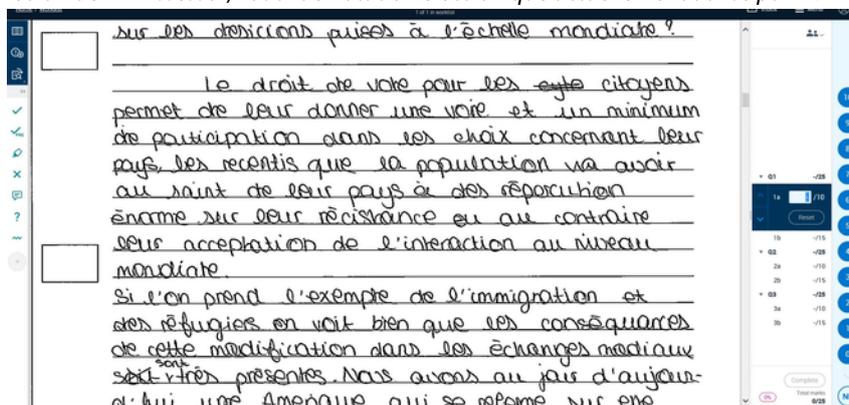
- La notation sur ordinateur permet d'améliorer la qualité de la notation des évaluations.
- Elle permet à l'IB de vérifier la notation des examinateurs « en temps réel ».
- Elle supprime les délais d'expédition et de réexpédition des copies d'examen ainsi que le risque de les perdre.
- Elle permet à l'IB de « sectionner » les copies des candidats et d'envoyer différentes questions à des correcteurs spécialisés (groupes d'éléments de question).
- Les examinateurs peuvent corriger le matériel audio et vidéo, les évaluations sur ordinateur ainsi que les copies manuscrites.

La notation sur ordinateur consiste à présenter aux examinateurs des images du travail des candidats au format électronique et à leur demander de noter les travaux directement sur ordinateur. Pour ce faire, ils utilisent un logiciel (« outil de notation »), qui leur permet de cocher certains éléments sur la version

électronique de la copie, de consigner les points attribués et d'insérer des remarques, exactement comme ils le feraient avec une version papier.

Figure 45

Capture d'écran de RM Assessor, l'outil de notation électronique actuellement utilisé par l'IB © RM Results



Les notes étant directement enregistrées par l'ordinateur, il est possible de vérifier le travail des examinateurs au fur et à mesure de leur notation. Disposer des copies d'examen au format électronique présente également des avantages considérables.

Premièrement, les copies peuvent être transmises rapidement aux examinateurs, et ce, en supprimant la nécessité d'attendre la réception de la version papier par voie postale et de deviner le nombre de copies que chacun pourra corriger. Les examinateurs reçoivent désormais les copies au moment où ils en ont besoin. Disposer d'un exemplaire des travaux des candidats directement dans le système constitue un avantage de taille, car cela évite tout dommage et toute perte.

Deuxièmement, cela permet d'inclure des « copies-tests », connues sous le nom de « copies de contrôle », dans les copies à corriger par les examinateurs, de façon à vérifier qu'ils respectent systématiquement la même norme tout au long de leur travail de notation.

Troisièmement, cela permet à l'IB de « découper » les copies d'examen en questions individuelles (connues sous le nom de « groupes d'éléments de question ») et de demander aux examinateurs de se concentrer sur la notation de ces questions plutôt que sur les épreuves intégrales.

Et, enfin, la notation sur ordinateur permet d'utiliser du matériel audio et vidéo. Cela ouvre considérablement l'éventail des travaux de candidats qui peuvent être envoyés comme preuves (par exemple, des enregistrements des représentations ou des épreuves orales).

Les différents types de barèmes de notation

Les barèmes de notation analytiques

Les barèmes de notation analytiques sont conçus pour les questions d'examen qui conduisent à une gamme étroite de réponses attendues des candidats.

Ces barèmes peuvent donner aux examinateurs des instructions spécifiques sur la manière de répartir le total des points prévus pour une question entre les différentes parties de la réponse. Même avec les questions structurées qui appellent des réponses hautement spécifiques, les barèmes de notation doivent fournir aux examinateurs des informations suffisantes pour qu'ils puissent noter de manière cohérente les principaux types d'approches différentes que les candidats pourraient adopter et les erreurs courantes qu'ils pourraient commettre. Les examinateurs devront inévitablement faire appel à leur jugement professionnel pour attribuer des points aux réponses inattendues ou aux autres réponses valides, mais les barèmes de notation devront leur apporter la meilleure orientation possible pour les aider à émettre un tel jugement.

Il est rarement possible de prévoir quelles parties d'une question les candidats réussiront ou non. Cette incertitude peut se révéler particulièrement problématique avec les questions structurées à réponse développée, dans lesquelles une erreur commise dans une des premières parties peut avoir des répercussions par la suite. L'IB conçoit ces questions de sorte que si un candidat commet une erreur dans la première partie de la question, il puisse poursuivre la question, mais les barèmes de notation analytiques doivent malgré tout apporter des orientations explicites quant à la manière de noter certains types de réponses incorrectes et de gérer le travail d'un candidat lorsqu'il a commis une erreur dans une partie d'une question.

Ce problème se pose tout particulièrement pour les matières comme les sciences ou les mathématiques, dans lesquelles l'IB attend des examinateurs qu'ils accordent des points en cas de réussite partielle. Sans structuration d'une question appelant une réponse détaillée, certains candidats risquent de ne pas réussir à obtenir beaucoup de points en raison d'une légère erreur commise au début de leur réponse ou d'un petit problème de compréhension de la question qui les conduit dans une mauvaise direction. En règle générale, les barèmes de notation analytiques les plus détaillés sont préparés pour les mathématiques.

Les critères d'évaluation

Lorsqu'une tâche d'évaluation est si ouverte que l'éventail des réponses envisageables est trop large pour que des barèmes de notation analytiques soient efficaces, il est possible de les remplacer par des critères (d'évaluation) descriptifs.

En principe, les critères d'évaluation ne font pas référence à un contenu spécifique attendu dans la réponse des candidats. Ils évoquent en revanche, dans la mesure du possible, la nécessité pour les candidats de montrer certaines connaissances du contenu. Les critères portent davantage sur le type d'accomplissement attendu des candidats, indépendamment des renseignements apportés dans chaque réponse. La gamme des niveaux de compétence possibles est illustrée par les descripteurs de niveaux.

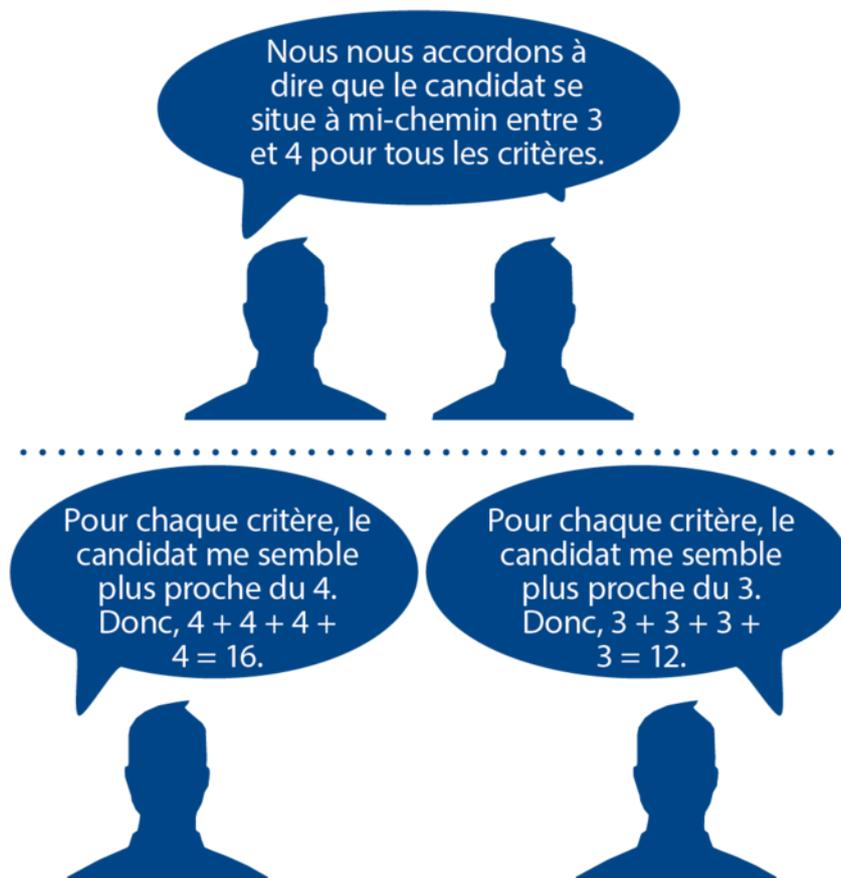
Une approche dite de meilleur ajustement est suivie par l'IB pour appliquer les descripteurs de niveaux de chaque critère. L'examineur qui utilise le critère d'évaluation doit choisir le niveau de réussite qui correspond le mieux au travail évalué. Il n'est pas nécessaire que chaque aspect du niveau de réussite soit satisfait dans le détail pour attribuer ce niveau. Il faut souligner que le plus haut niveau d'un critère, quel qu'il soit, ne décrit pas un travail parfait.

Même s'il est tout à fait possible de noter un travail en utilisant un seul critère, il est plus fréquent de le faire à partir de plusieurs critères. Les différentes notes sont ensuite additionnées, les critères considérés comme plus importants étant notés sur un nombre de points plus élevé. Dans de telles situations, il est important que les critères soient indépendants les uns des autres, car il ne serait pas juste qu'un candidat reçoive des points à plusieurs reprises pour le même aspect de sa réponse.

Utiliser plusieurs critères pose un autre problème : cela réduit souvent la fiabilité et la cohérence de la notation. Il est en effet plus probable que les examinateurs aient à décider plusieurs fois du « meilleur ajustement » et, même si leur désaccord ne porte que sur un point, celui-ci peut être multiplié si les différences s'ajoutent toutes dans le même sens.

Figure 46

Manière dont les critères d'évaluation peuvent amener des examinateurs partageant le même avis général sur la qualité du travail d'un candidat à attribuer des notes très différentes



Il est habituel de réutiliser les mêmes critères d'évaluation chaque année, même si les questions posées aux candidats sont différentes. Cela s'explique par le fait que ce qui est évalué reste de même nature ; toute nouvelle question porte généralement sur des détails spécifiques qui ne figurent pas explicitement dans les critères (contrairement aux barèmes de notation analytiques). L'examinateur principal préparera souvent des remarques sur la notation pour expliquer aux examinateurs comment appliquer les critères d'évaluation à chaque question et inclura éventuellement des exemples de détails pertinents. Lorsque les critères d'évaluation sont utilisés pour l'évaluation interne, les enseignants et les réviseurs de notation doivent se référer au matériel de soutien pédagogique publié, afin de consulter des exemples d'application des critères.

Bien que les descripteurs de niveaux de chaque critère soient par nature hiérarchiques et qu'ils contiennent souvent la hiérarchie des compétences cognitives définies par Bloom *et al.* (1956), il est important de garder à l'esprit qu'il n'existe pas de lien direct entre les exigences cognitives et les niveaux des critères. Les descripteurs des niveaux les plus bas ne sont pas uniquement consacrés aux compétences cognitives « plus simples », et les descripteurs des niveaux les plus élevés ne sont pas uniquement réservés pour les compétences cognitives « supérieures ». Il doit exister toute une gamme de niveaux pour chaque domaine de compétence cognitive décrit par Bloom.

Les critères globaux – Les bandes de notation

Il est parfois inapproprié de séparer les différents critères d'évaluation pour noter un travail. Cela se produit généralement lorsqu'il est impossible d'avoir des critères distincts, indépendants les uns des autres. Le cas

échéant, des bandes de notation viennent remplacer les critères distincts. De fait, les bandes de notation constituent ensemble un critère global appliqué au travail, qui est jugé comme un tout. Étant donné la nécessité de disposer d'une gamme de notes raisonnable pour pouvoir différencier les accomplissements des candidats, plusieurs notes correspondent à chaque descripteur de niveaux.

Figure 47

Exemple de bandes de notation pour l'épreuve 1 (NM et NS) figurant dans le Guide d'histoire du Programme du diplôme publié en 2015

Points	Descripteurs de niveaux
0	La réponse n'atteint pas l'un des niveaux décrits ci-dessous.
1 – 3	La réponse se concentre insuffisamment sur la question. Les sources sont mentionnées, mais à ce niveau, les références sont le plus souvent des descriptions du contenu des sources plutôt que des éléments utilisés en tant que preuves pour étayer l'analyse. Aucune connaissance personnelle n'est utilisée ou, lorsque des connaissances personnelles sont présentes, elles sont erronées ou hors sujet.
4 – 6	La réponse se concentre généralement sur la question. Les sources sont mentionnées, et ces références sont utilisées en tant que preuves pour étayer l'analyse. Lorsque des connaissances personnelles sont utilisées, elles manquent de pertinence ou d'exactitude. Aucun ou peu d'efforts sont faits pour synthétiser les connaissances personnelles et le contenu des sources.
7 – 9	La réponse se concentre sur la question. Les sources sont clairement mentionnées et ces références sont efficacement utilisées en tant que preuves pour étayer l'analyse. Des connaissances personnelles exactes et pertinentes sont utilisées. Une synthèse efficace des connaissances personnelles et du contenu des sources est manifeste.

Les descripteurs eux-mêmes ont tendance à être assez longs puisqu'ils couvrent tout un éventail de qualités pouvant être présentes dans les travaux des candidats, et ils ont un lien direct avec les objectifs du cours. Tout comme pour les critères d'évaluation, une approche dite de meilleur ajustement est utilisée. En outre, les examinateurs doivent émettre un jugement en choisissant une note dans la gamme de notes disponibles dans chaque descripteur de niveaux. Pour ce faire, ils déterminent à quel point le travail du candidat correspond à ce descripteur. Par exemple, un descripteur de niveaux peut couvrir les notes 6 à 10. L'examinateur attribuera une note comprise entre 6 et 10 en fonction de la mesure dans laquelle le travail du candidat correspond à ce descripteur de la bande de notation.

Les recherches ont montré que, lorsque des critères d'évaluation et une évaluation globale (bande de notation) sont utilisés pour noter une dissertation se prêtant à une évaluation à l'aide de ces deux méthodes, il existe peu de différences entre ces dernières en ce qui concerne la fiabilité de la notation (Wood, 1991).

Un soutien supplémentaire pour les examinateurs

Les examinateurs, comme tous les apprenants, tirent profit des multiples manières d'assimiler les informations qui leur sont proposées.

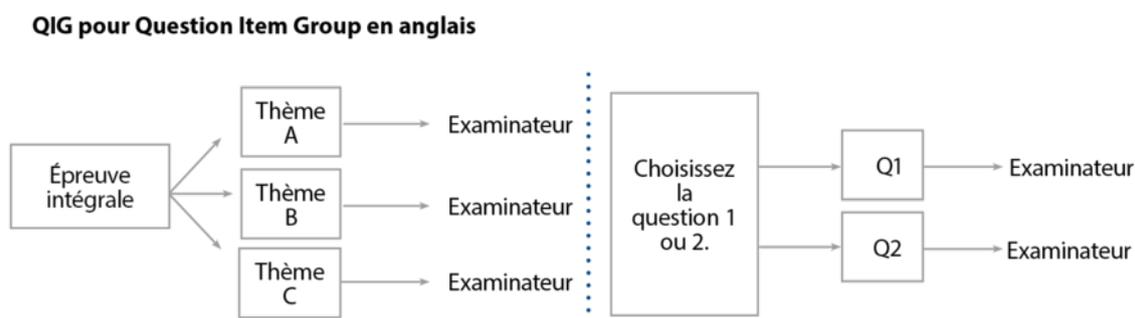
En plus des barèmes de notation et des critères d'évaluation au format papier, l'IB utilise de plus en plus des vidéos pour expliquer à la fois oralement et visuellement la norme de notation attendue. Par ailleurs, l'IB encourage les examinateurs à poser leurs questions et l'équipe des examinateurs superviseurs à leur faire un retour d'information par téléphone, par messagerie grâce au logiciel de notation électronique ou par courriel.

La communication est intégrée dans le modèle d'assurance de la qualité de l'IB. En effet, lorsqu'un examinateur montre dans une copie de contrôle ou de validation qu'il ne respecte pas la norme attendue, son chef d'équipe doit impérativement lui faire un retour d'information. Si sa notation est conforme à la norme attendue, il a tout de même la possibilité de bénéficier d'un retour d'information.

Les groupes d'éléments de question (QIG pour Question Item Group en anglais)

Figure 48

Les deux façons de diviser une épreuve en groupes d'éléments de question



La création des QIG repose sur l'idée qu'il est plus facile pour les examinateurs de corriger plusieurs fois la même question que de corriger toutes les questions d'une copie avant de recommencer avec la première question pour le candidat suivant.

Il existe deux façons de diviser une épreuve en QIG. La première consiste à suivre les questions ou les thèmes. Dans ce cas de figure, il est important de garder dans le même QIG les parties des questions qui sont liées. Par exemple, si la partie (a) conduit à la partie (b), toutes deux doivent figurer dans le même QIG. Les examinateurs peuvent ainsi attribuer des points de suivi mais aussi attribuer des points lorsqu'un candidat a indiqué dans la première partie de sa réponse des éléments pertinents pour la seconde partie.

En revanche, plus la quantité de détails du barème de notation à mémoriser par l'examineur est réduite, plus son travail sera cohérent. Par conséquent, plusieurs petits QIG sont préférables à un grand. Cet aspect est pris en considération pour décider de la division des épreuves d'examen.

La seconde façon de diviser une épreuve en QIG intervient lorsque le candidat doit répondre à une question de son choix. Dans ce cas, chaque question possible devient un QIG, qui représente l'intégralité de la copie du candidat pour un examen. Par exemple, imaginez qu'un candidat ait le choix entre une question portant sur la poésie ou une question portant sur la prose pour une dissertation de littérature. La question sur la poésie devient alors le QIG 1, et l'examineur spécialisé en poésie peut corriger toutes les copies d'examen concernées. De la même manière, la question sur la prose devient le QIG 2, et l'examineur spécialisé en prose peut corriger toutes les copies d'examen portant sur cette question.

Pour chaque QIG, il existera des copies d'entraînement, de validation et de contrôle particulières, et les examinateurs devront prouver qu'ils sont capables de noter conformément à la norme attendue pour leur QIG. Même si cela peut paraître difficile, cette mesure permet en réalité à un examinateur n'ayant pas compris la norme pour une question précise d'une épreuve de corriger les autres questions, au lieu d'être suspendu de la notation parce qu'il ne peut pas noter l'épreuve dans son intégralité conformément à la norme.

Le modèle d'assurance de la qualité

L'IB procède à présent à la notation électronique de presque tous les travaux évalués en externe. Le travail des examinateurs est vérifié grâce à un système qui exige d'eux et de l'examineur principal qu'ils notent les mêmes copies et qui compare ensuite les notes des examinateurs à celles de l'examineur principal pour la composante concernée. Ce processus fonctionne de la manière suivante.

1. L'examineur principal ainsi que quelques membres de son équipe d'examineurs superviseurs notent un certain nombre de copies au cours de la semaine qui suit la fin des examens, dans le cadre du processus de normalisation.
2. Les notes définitives attribuées par l'équipe d'examineurs superviseurs sont enregistrées dans le système avec les remarques ou annotations correspondantes.
3. L'IB détermine, en consultation avec l'équipe d'examineurs superviseurs, la marge acceptable entre la note définitive (la note accordée par l'examineur principal) et la note accordée par les examinateurs à ces copies. Cette marge acceptable porte le nom de « marge de tolérance ».
4. Les remarques et les notes sont ensuite masquées, et les copies sont replacées dans la série de copies que les examinateurs devront noter.
5. Les copies comportant les notes définitives sont utilisées pour l'entraînement à la notation, le processus de normalisation et le processus de contrôle.

Le modèle d'assurance de la qualité repose sur l'idée que les examinateurs ont la possibilité de tirer des enseignements du travail réalisé tout au long du processus de notation. Ainsi, lorsque les copies de contrôle sont notées en dehors de la marge de tolérance, les examinateurs peuvent tout de suite consulter les notes définitives, ce qui leur permet de comprendre qu'ils n'ont pas appliqué le barème de notation de façon appropriée et de voir la marche à suivre pour améliorer leur travail. Les copies de contrôle que les examinateurs ont notées en respectant la marge de tolérance définie leur sont également régulièrement présentées, afin qu'ils puissent s'approcher autant que possible de la note définitive.

Les examinateurs doivent donc considérer les copies de contrôle comme des possibilités de perfectionnement professionnel. La plupart des examinateurs recevront un retour d'information sur leur notation des copies de contrôle à un moment ou un autre du processus de notation. Ce retour d'information doit être apprécié et non constituer une source d'inquiétude.

Les copies d'entraînement

- L'objectif des copies d'entraînement est d'aider les examinateurs dans leur apprentissage de la norme de notation.
- Ces exemples doivent montrer comment noter les travaux de candidats typiques ou toute situation courante dans laquelle les examinateurs comprennent mal le barème de notation.

Les exemples de travaux de candidats comportant les notes définitives sont utilisés par les examinateurs pour vérifier leur compréhension du barème de notation et des remarques sur la notation. Une fois ces copies passées en revue, les examinateurs devraient avoir la certitude qu'ils peuvent procéder à une notation correcte.

Lors de la sélection des copies d'entraînement, les examinateurs principaux doivent tenir compte des éléments suivants :

- l'objectif étant d'aider les examinateurs à se préparer au processus de notation, il est inutile d'inclure des copies d'examen qui illustrent une situation qu'ils ne sont pas susceptibles de rencontrer à nouveau ;
- la nécessité d'expliquer les notes attribuées, à l'aide de commentaires informatifs ;
- l'utilité ou la pertinence d'exclure certaines parties de la copie d'examen d'un candidat ou d'ajouter une remarque indiquant qu'elles ne sont pas utiles à l'apprentissage des examinateurs.

Pour être de qualité, une série de copies d'entraînement doit inclure les éléments suivants (par ordre de priorité) :

1. des exemples montrant comment noter un travail typique ;
2. toutes les situations courantes dans lesquelles les examinateurs comprennent mal le barème de notation ;
3. une large gamme de notes, allant de faibles à élevées, pour aider les examinateurs à reconnaître où ils peuvent attribuer des points et comment distinguer les bonnes (et les mauvaises) réponses ;
4. des exemples d'exceptions et de réponses complexes, et comment les traiter.

Les copies de validation

- L'objectif des copies de validation est de montrer ou de prouver que les examinateurs notent selon la norme de notation adoptée.
- Ces copies comprennent des exemples des réponses les plus courantes données par les candidats sur l'ensemble de l'éventail de notes.

Avant d'être autorisés à commencer à noter des travaux de candidats « en conditions réelles », les examinateurs doivent montrer qu'ils sont capables de noter selon la norme adoptée. Pour ce faire, ils sont « évalués » à l'aide de cinq exemples de travaux de candidats déjà notés par l'examineur principal, ce qui permet de savoir au préalable quelle note ils sont censés attribuer.

Si la note attribuée est différente de celle de l'examineur principal, un retour d'information est fait aux examinateurs pour leur indiquer la note correcte et les raisons de cette attribution. Après avoir réfléchi sur ces éléments, les examinateurs se voient offrir une seconde possibilité de montrer qu'ils comprennent à présent la norme de notation.

Pour être de qualité, une série de copies de validation doit inclure les éléments suivants (par ordre de priorité) :

1. des exemples des réponses les plus courantes données par les candidats sur l'ensemble de l'éventail de notes (élevées, moyennes et faibles) ;
2. des exemples de travaux exigeant de l'examineur qu'il ait compris toutes les exceptions courantes du barème de notation ou les erreurs courantes que les candidats sont susceptibles de faire.

Lors de la sélection des copies de validation, les examinateurs principaux ne doivent pas inclure des copies d'examen destinées à « piéger » les examinateurs. Ils doivent choisir des copies difficiles à noter, qui sont représentatives du type de difficultés auxquelles les examinateurs seront confrontés de façon quotidienne. Ils doivent également chercher des exemples de tous les cas de figure importants que les examinateurs doivent être en mesure de traiter correctement.

Les copies de contrôle

- L'objectif des copies de contrôle est de montrer ou de prouver que les examinateurs continuent de noter selon la norme adoptée.
- Ces copies sont représentatives des erreurs courantes que peuvent faire les examinateurs et contiennent des exemples de réponses sur l'ensemble de l'éventail de notes.

Même si un examinateur ne peut pas être autorisé à commencer à noter des travaux de candidats en conditions réelles avant d'avoir compris la norme de notation, il est évident que sa manière de noter peut évoluer au fil du temps. Par conséquent, sa notation est régulièrement vérifiée avec des copies de contrôle.

Les copies de contrôle ont déjà été notées par l'examineur principal, ce qui permet de savoir au préalable quelles notes sont censées être attribuées. Ces copies ressemblent à tous les autres travaux des candidats. L'examineur ne sait donc pas qu'il corrige une copie de contrôle.

Si l'examineur donne la note de l'examineur principal à une copie de contrôle, il peut poursuivre son travail de notation. Toutefois, s'il montre qu'il ne respecte plus la norme de notation attendue, une intervention est nécessaire.

Cela commence par un retour d'information et des orientations, afin de laisser la possibilité à l'examineur de retrouver la norme de notation. Si tel est le cas, l'examineur peut ensuite reprendre sa notation, sachant que tous les candidats pourront obtenir une note cohérente, quelle que soit la personne qui corrige leur travail.

Si l'examineur est incapable de retrouver la norme de notation attendue, il ne pourra plus corriger de travaux de candidats.

En règle générale, les examinateurs doivent corriger une copie de contrôle après dix copies notées en conditions réelles. Étant donné qu'il ne s'agit pas forcément de la dixième copie, les examinateurs ne savent pas à quel moment ils corrigent une copie de contrôle. Si la situation l'exige, la fréquence des copies de contrôle peut être modifiée.

Pour être de qualité, une série de copies de contrôle doit inclure les éléments suivants (par ordre de priorité) :

1. des travaux de candidats représentatifs des erreurs courantes que peuvent faire les examinateurs ;
2. une large gamme de notes (élevées, moyennes et faibles).

Contrairement aux copies de validation, les copies de contrôle sont réparties de façon aléatoire. Chaque examinateur les recevra donc dans un ordre différent.

Lors de la sélection des copies de contrôle, les examinateurs principaux ont bien conscience qu'à ce stade, les examinateurs corrigent des travaux de candidats en conditions réelles. Par conséquent, l'important est de s'assurer que les notes attribuées aux candidats sont correctes. Il est donc approprié d'inclure des copies complexes ou difficiles à noter. Cela n'est pas effectué dans le but de « piéger » les examinateurs, mais de s'assurer que ces derniers respectent la norme de notation. Il est également possible d'utiliser des copies de contrôle afin de vérifier que les examinateurs suivent toutes les règles de notation, et pas uniquement que leur notation est correcte. Même si cette situation est rare, il peut arriver que l'examineur principal choisisse une copie de contrôle qui doit remonter jusqu'au chef d'équipe pour une raison ou une autre (par exemple, des preuves d'une mauvaise conduite ou une réponse totalement inattendue).

Les chefs d'équipe doivent superviser les examinateurs et leur apporter leur soutien tout au long de la notation. Dès qu'un examinateur est contraint d'arrêter de noter, car une de ses copies de validation ou de contrôle a été notée de façon incorrecte, son chef d'équipe doit se mettre en rapport avec lui pour lui faire un retour d'information et le conseiller. Le chef d'équipe et l'examineur pourront consulter la copie de contrôle qui n'a pas été notée selon la norme, ainsi que la note attribuée par l'examineur principal et ses annotations. Le rôle du chef d'équipe est d'expliquer à l'examineur pourquoi la note attribuée n'était pas correcte. Ils pourront ensuite discuter ensemble du barème de notation et de la bonne façon de l'appliquer. Les examinateurs qui continueront à appliquer le barème de notation de façon incorrecte, malgré la formation complémentaire assurée par le chef d'équipe, seront définitivement déchargés de leur travail de notation.

La marge de tolérance

- La marge de tolérance reflète les différences légitimes qui peuvent exister entre les notes attribuées à un même travail par des examinateurs différents. Imaginez, par exemple, que deux enseignants de votre établissement corrigent le même travail. Ils conviennent tous deux de la bonne qualité du travail, mais le premier attribue 46 points au travail et le second 47 points.
- L'IB s'engage à poser des questions qui évaluent ce qui est important, et pas simplement ce qui est facile à corriger.

Pour les questions appelant une réponse qui est soit bonne soit mauvaise, l'examineur doit attribuer exactement la même note que l'examineur principal. Pour les autres questions, notamment celles qui évaluent la compréhension ou l'analyse, un certain degré de jugement intervient dans la note qui sera

attribuée. Deux examinateurs spécialisés ayant la même conception d'une bonne réponse attribueront des notes proches, qui pourront toutefois varier d'un ou deux points.

L'idée de marge de tolérance est utilisée pour rendre compte de ces différences d'opinion légitimes lors de l'examen du travail de notation des examinateurs. Par exemple, si l'examineur principal a attribué une note de 46 à une dissertation et qu'une marge de tolérance de 2 points était associée à la question, alors il est considéré que tout examinateur attribuant une note comprise entre 44 et 48 respecte la norme de notation.

Pour les questions en plusieurs parties ou notées à l'aide de plusieurs critères, il convient de surveiller les différences au niveau de la note globale (c'est-à-dire totale) mais aussi au niveau de chaque partie ou critère. Un examinateur doit se trouver dans la marge de tolérance pour ces deux aspects pour montrer qu'il a respecté la norme de notation.

Les copies difficiles ou inhabituelles

Si les examinateurs rencontrent des copies qu'ils ne savent pas comment noter, leur premier réflexe doit être de demander conseil à leur chef d'équipe. Si malgré cela, ils estiment qu'ils ne peuvent toujours pas garantir une notation équitable, ils peuvent demander à un examinateur placé plus haut dans la hiérarchie de noter ces copies. Les copies particulièrement problématiques seront alors portées à l'attention de l'examineur principal, qui prendra une décision définitive quant à la manière de les noter.

De la même manière, toutes les copies inhabituelles et les copies des candidats ayant répondu à des épreuves d'examen modifiées seront corrigées par l'examineur principal, qui devra veiller de près à ce que la norme de notation pour ces copies soit équivalente à celle appliquée pour toutes les autres.

Cette notation aura lieu une fois les seuils d'attribution des notes finales fixés, afin que les examinateurs superviseurs aient la même compréhension des notes finales pour cette session d'évaluation particulière. Même si une note est attribuée, dans ces cas spécifiques, il est demandé aux examinateurs de tenir compte des seuils d'attribution des notes finales pour s'assurer que les normes de notation spécifiquement établies pour ces candidats sont adaptées aux seuils d'attribution des notes finales fixés pour la cohorte globale.

Liens avec un établissement scolaire

L'IB part du principe que les examinateurs ne doivent pas noter le travail de leurs candidats.

À cet effet, les examinateurs sont tenus d'informer l'IB de tout lien avec des candidats et des établissements scolaires, afin qu'il ne leur soit pas demandé de noter leurs travaux. La division de l'évaluation de l'IB dispose également d'une politique en matière de conflit d'intérêts, qui concerne les personnes travaillant pour l'IB susceptibles d'avoir des liens avec des établissements scolaires.

Si vous avez des raisons de penser qu'un examinateur ou un membre du personnel de l'IB n'a pas déclaré ses liens avec un établissement scolaire ou un candidat, veuillez envoyer un courriel à complaints@ibo.org ou vous mettre en rapport avec le service [L'IB vous répond](#).

Les commentaires des examinateurs

On ne peut pas s'occuper de deux choses à la fois. C'est comme les citrouilles dans l'eau : l'une refait surface tandis que l'on essaie de maintenir l'autre sous l'eau.

(Proverbe chinois)

L'évaluation sommative de l'IB a pour objectif de mesurer les accomplissements des candidats. Il est donc demandé aux examinateurs de se concentrer exclusivement sur la notation du travail des candidats, en respectant la norme requise. Ils ne doivent inclure des commentaires que si cela les aide dans leur travail de notation.

Ils ne doivent inclure des commentaires que si cela les aide dans leur travail de notation. Pour rédiger des commentaires constructifs à l'intention des candidats ou des enseignants, l'examineur doit déterminer la note qui correspond au travail, puis tenter d'expliquer de quelle manière ce travail aurait pu être amélioré. La seconde tâche n'est pas insignifiante, car elle constitue l'une des caractéristiques essentielles d'un bon

enseignement. Elle requiert cependant du temps et une réflexion qui éloigneront l'examineur de sa mission principale, à savoir corriger le travail selon une norme de notation cohérente. En termes simples, l'IB attend des examinateurs qu'ils effectuent une seule tâche (notation) en maintenant un niveau élevé de qualité, plutôt que deux tâches (notation et retour d'information) avec un niveau de qualité moindre.

L'examineur ne peut émettre un jugement qu'à partir du travail mis à sa disposition, alors que les enseignants expérimentés pourront se baser sur tout un éventail d'informations pour décider du retour d'information à faire au candidat. Le retour de l'examineur sera donc moins complet que celui de l'enseignant.

Pour toutes ces raisons, l'IB indique très clairement à ses examinateurs qu'ils doivent uniquement noter le travail du candidat en respectant la norme, et non ajouter des commentaires pour faire un retour d'information au candidat ou à l'enseignant.

Les examinateurs doivent indiquer clairement à quels endroits les points ont été attribués et, en cas d'ambiguïté, expliquer ce choix en ajoutant des commentaires appropriés. Cela permet à l'IB de vérifier les normes de notation et rend le processus d'attribution des points plus transparent pour les établissements.

Le rapport de recorection de catégorie 1, qui peut être demandé par les établissements à l'occasion d'une réclamation concernant les résultats, constitue la seule exception à ce principe de rédaction des commentaires pour expliquer la notation uniquement. Dans ce cas précis, un examinateur haut placé dans la hiérarchie répondra aux préoccupations exprimées par l'établissement dans le cadre de la réclamation concernant les résultats, et ce, avec un niveau de détail plus poussé que ce qui est généralement attendu des examinateurs.

Totalisation

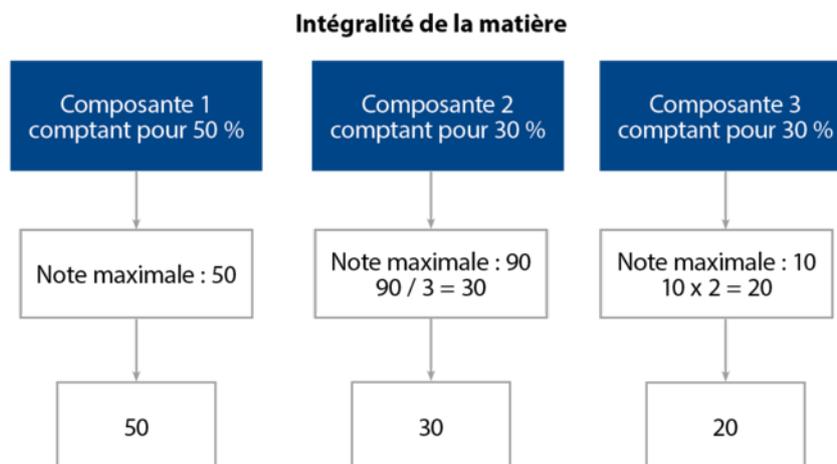
- Le terme « totalisation » désigne le processus d'addition des notes obtenues dans les composantes pour produire un résultat général.
- Pour que chaque composante contribue à la note définitive dans les bonnes proportions (pondération), il peut s'avérer nécessaire d'appliquer un coefficient aux notes des composantes.
- L'IB utilise un « système de compensation » grâce auquel les candidats peuvent compenser de mauvais résultats obtenus dans une question ou une composante par un niveau plus élevé atteint dans une autre.
- La note finale par matière d'un candidat est calculée en totalisant les notes obtenues dans chaque composante, et non ses notes finales par composante.

Le terme « totalisation » désigne le processus d'addition des notes (et des seuils) issues des différentes composantes pour produire une note définitive ou un seuil d'attribution des notes finales global. Pour y parvenir, il peut s'avérer nécessaire d'appliquer un coefficient aux notes globales (ou seuils) des composantes.

Un coefficient est appliqué afin de conserver la pondération voulue pour chaque composante pour ce qui est de sa contribution à l'évaluation globale du cours. Cela implique de multiplier ou de diviser les notes des composantes pour qu'elles contribuent comme il se doit au total global de la matière. Cela vaut également pour les seuils d'attribution des notes finales fixés pour la composante, qui auront initialement été déterminés à partir du nombre de points maximum pour les composantes.

Figure 59

Exemple d'application d'un coefficient pour une matière comportant trois composantes



Le concept de pondération permet de refléter l'importance relative que l'IB accorde aux éléments évalués, en fonction de la contribution au résultat final qu'il souhaite pour chacun. Ainsi, si une composante sert avant tout à évaluer l'interprétation des données ou des sources et que sa pondération est de 30 %, cela implique que, en comparaison des autres objectifs spécifiques du cours, l'interprétation représente 30 % de ce qui est important. Il est fréquent que plusieurs composantes évaluent les mêmes objectifs spécifiques, ce qui rend ce calcul moins pertinent.

Un aspect secondaire, mais important, de la pondération, est qu'elle permet de déterminer le nombre total de points d'une évaluation qui convient aux tâches et aux critères de notation, plutôt que d'essayer de créer artificiellement un total global.

Il est très important de souligner que les notes d'un candidat n'ont pas à correspondre à la pondération de la composante. L'IB est conscient que les candidats n'ont pas tous les mêmes points forts ; c'est pourquoi il utilise des outils d'évaluation très divers. Il est important de comparer les pondérations réelles des résultats de la cohorte entière à celles indiquées lors de la conception de la matière. Si, au cours d'une session, la contribution d'une composante est beaucoup plus forte que prévu, cela peut indiquer que l'épreuve était particulièrement facile et donc qu'il faut chercher d'autres preuves pour l'attribution des notes finales. En revanche, si la contribution d'une composante est régulièrement beaucoup plus forte que prévu, un réexamen de la conception est nécessaire, car cela indique que l'évaluation ne fonctionne pas comme prévu.

L'approche présentée ne reflète pas les méthodes plus complexes de pondération, de calcul (totalisation) et d'application d'un coefficient décrites, par exemple, par Wood (1991, chapitre 10), mais elle est basée sur les principes solides de l'évaluation critériée et soutient les principes de l'évaluation de l'IB.

Cette approche d'obtention de la note définitive par totalisation implique qu'un candidat peut compenser de mauvais résultats obtenus à une composante par des résultats plus élevés obtenus dans une autre, car seule la note totale a des conséquences sur la note finale. Il s'agit là de la notion incarnée par le « système de compensation », par opposition au « système de maîtrise » qui impose au candidat d'afficher le niveau requis dans toutes les composantes pour se voir attribuer une note finale particulière.

Il convient de souligner que la note finale d'un candidat à une matière est obtenue par la totalisation des notes des différentes composantes, et non des notes finales des composantes. Étant donné que chaque note finale d'une composante représente une gamme de notes, il est fort possible que deux candidats avec les mêmes notes finales par composante se voient attribuer des notes finales différentes pour la matière.

Révision de notation

- La révision de notation consiste à vérifier la norme de notation des enseignants et non à recorriger le travail des candidats.
- Une révision de notation réussie signifie que les candidats obtiendraient la même note d'évaluation interne s'ils étudiaient dans un autre établissement situé à l'autre bout du monde. Cela fait référence à la notion de norme globale.
- Parmi les preuves analysées pendant la révision de notation figurent les explications fournies par les enseignants pour justifier leurs notes, et pas uniquement la qualité du travail des candidats.

Grâce à l'échantillonnage dynamique (la nouvelle approche de la révision de notation), l'IB est certain que chaque réviseur de notation respecte la norme de l'examineur principal, et cela réduit la révision à une seule phase.

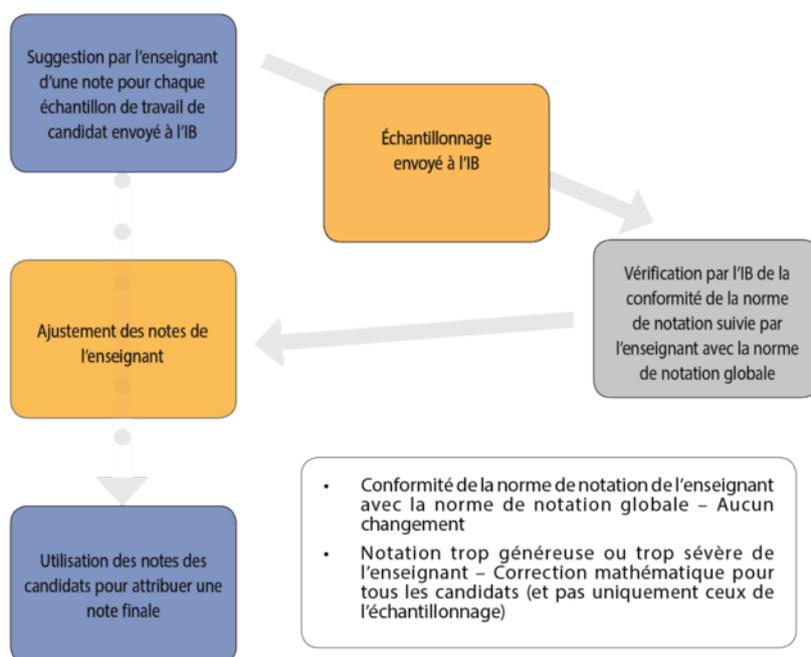
Qu'est-ce que la révision de notation ?

Bien souvent, il est difficile de mesurer la caractéristique souhaitée en évaluant le candidat à l'aide d'un examen officiel limité dans le temps. L'approche la plus valide consiste donc généralement à demander à l'enseignant de s'en charger à l'aide d'une évaluation interne. La section intitulée « [Le rôle de l'évaluation se déroulant en classe et de l'évaluation interne](#) » explique en quoi une telle approche est adaptée.

Malgré les résultats importants qu'elle produit, cette approche s'accompagne du risque d'une différence d'interprétation de la norme de notation par les enseignants. Ainsi, deux enseignants dans deux établissements différents pourraient attribuer deux notes différentes à un même travail. Pour que tous les examinateurs comprennent de la même façon la norme attendue, l'IB s'appuie sur un processus complet de formation et de test, mais il ne peut pas faire de même avec tous les enseignants.

Au lieu de cela, il croit en la capacité des enseignants à noter tous les travaux avec cohérence ; c'est pourquoi il n'a plus qu'à appliquer un facteur arithmétique si cela est nécessaire pour ramener leur notation au niveau de la norme globale. Pour ce faire, il demande aux enseignants de fournir un échantillonnage de leur notation, qui sera comparé à la norme de l'examineur principal. À partir des données produites par cette comparaison, une formule mathématique est calculée, si nécessaire, pour ajuster toutes les notes fournies par chaque enseignant.

Figure 49
Présentation de la révision de notation



Le diagramme ci-dessus explique comment l'échantillonnage est révisé par l'examinateur. Sur la base d'une comparaison statistique entre les deux séries de notes, un ajustement est apporté aux notes attribuées par l'enseignant (pour cette composante) à tous les candidats de l'établissement, si cela est nécessaire. Si les notes attribuées par l'enseignant sont systématiquement trop basses ou trop élevées, l'ajustement sera le même pour chacune des notes qu'il a attribuées. Toutefois, si les notes attribuées par l'enseignant sont trop basses ou trop élevées uniquement dans la tranche de notes supérieure ou la tranche de notes inférieure, l'ajustement pourra varier dans toute la gamme de notes.

Points importants à retenir

- La révision de notation vise à vérifier le degré de précision et de cohérence avec lequel l'enseignant a appliqué les critères d'évaluation lors de la notation des travaux.
- À l'issue de la révision de notation, les notes de l'établissement peuvent être augmentées ou diminuées, ou elles peuvent rester inchangées.
- Si un facteur de révision de notation est appliqué, cela ne signifie pas que la notation de l'enseignant est mauvaise, mais simplement qu'il s'est montré plus ou moins généreux que la norme globale.

Pour des raisons pratiques, l'IB révisé la notation des établissements scolaires et non celle des enseignants. Il est donc fondamental que tous les enseignants d'un même établissement veillent à noter les travaux en respectant la même norme, afin que l'équité soit au rendez-vous lorsque l'IB applique un facteur de révision de notation aux notes de tous les enseignants.

Figure 50

Éléments à prendre en considération par le coordonnateur lors de l'analyse des notes d'évaluation interne destinées à la révision de notation

<p>Dans l'établissement, tous les enseignants* suivent-ils la même norme de notation ?</p> <p>Les enseignants notent-ils constamment tous les travaux des candidats en suivant la même norme ?</p> <p>Les enseignants expliquent-ils clairement les raisons de l'attribution des notes ?</p> <p>* Les enseignants chargés d'attribuer des notes d'évaluation interne.</p>	
---	--

Pour de plus amples informations sur les aspects techniques du calcul des facteurs de révision de notation, veuillez vous reporter à l'annexe 1 « Révision de notation de l'évaluation interne ».

La sélection des travaux de candidats

- Tous les candidats doivent recevoir une note juste et, de ce fait, le travail de n'importe quel candidat doit pouvoir illustrer la norme suivie par l'enseignant et servir d'exemple pour la révision de notation.
- Deux enseignants peuvent attendre différentes choses d'un très bon ou d'un très mauvais travail ; il est donc essentiel que les travaux inclus dans l'échantillonnage représentent tout l'éventail de notes.
- Pour garantir la transparence et l'équité entre tous les établissements scolaires, l'IB doit pouvoir consulter des preuves montrant le travail à la base de la note définitive d'un candidat.

Pour garantir la transparence et éviter toute impression de mauvaise conduite, il est important que ce soit l'IB, et non l'établissement scolaire, qui sélectionne les travaux à inclure dans l'échantillonnage destiné à la révision de notation. L'IB suit des principes généraux (voir ci-dessous) pour s'assurer que le facteur de révision de notation est le plus fiable possible et, dans la limite de ces principes, que les candidats sont sélectionnés de manière aléatoire.

Le premier principe est que l'IB doit utiliser le plus petit échantillonnage possible pour obtenir un facteur de révision de notation fiable, quel que soit l'établissement scolaire concerné. Cette mesure réduit la charge qui repose sur l'établissement, mais aussi les coûts que l'IB pourrait avoir à répercuter sur les établissements en les incluant dans les frais d'examen. Il est légitime que le nombre de travaux inclus dans l'échantillonnage soit différent en fonction des établissements si l'IB a besoin de plus d'exemples de la notation des enseignants pour déterminer un facteur de révision de notation solide. C'est pourquoi des échantillons supplémentaires sont parfois réclamés. Étant donné qu'une telle demande peut survenir lors du processus de révision de notation, tous les travaux de candidats doivent être disponibles jusqu'à la publication des résultats.

Le second principe est qu'il faut être certain que le facteur de révision de notation est juste pour tous les candidats, quelle que soit leur note. L'expérience a prouvé que les enseignants pouvaient avoir des attentes différentes à plusieurs niveaux de la gamme de notes et qu'un enseignant dont la notation est plus généreuse que la norme globale pour un travail de mauvaise qualité pouvait aussi être plus sévère que la norme globale pour un travail de bonne qualité. L'échantillonnage de travaux d'évaluation interne est donc soigneusement sélectionné de manière à assurer une représentation appropriée de toute la gamme de notes de l'établissement.

L'IB tend à ne pas sélectionner les candidats ayant obtenu la note maximale pour la révision de notation. Cette exclusion vise l'équité, car elle permet d'offrir aux candidats de la tranche de notes supérieure la possibilité d'obtenir une meilleure note à la suite d'une révision de notation, si l'enseignant s'est montré trop sévère dans sa notation. De même, l'IB ne sélectionne habituellement pas les candidats ayant obtenu un zéro.

Les travaux inhabituels et « atypiques »

Il est important que chaque candidat soit noté de façon équitable par l'enseignant. Par conséquent, pour chaque candidat, les notes et les remarques de l'enseignant devraient être représentatives de la norme suivie. Cela signifie que, sauf exception, un établissement scolaire devrait toujours pouvoir envoyer un travail à l'IB si cela lui était demandé pour la révision de notation.

La révision de notation a pour but de vérifier que les enseignants notent les travaux en respectant la norme globale, et ce, même si un travail est particulièrement difficile à noter. Le cas échéant, l'enseignant peut avoir à justifier avec plus de détails la note attribuée. Par exemple, si le candidat a bénéficié d'une aide supplémentaire, l'enseignant doit le signaler dans sa notation et l'examineur doit prendre cet élément en compte lorsqu'il révisé la notation de l'enseignant pour ce travail.

L'impossibilité de trouver un facteur de révision de notation

Dans certains cas, il est impossible de calculer un facteur de révision de notation sur la base de l'échantillonnage de travaux reçu. Cela se produit lorsque la différence entre la notation de l'examineur et celle de l'enseignant n'est pas cohérente, ou que l'examineur estime que l'enseignant a été trop généreux ou au contraire trop sévère par rapport à la norme globale. Le cas échéant, l'IB demande à l'établissement d'envoyer des travaux supplémentaires afin de pouvoir garantir l'application d'un facteur de révision de notation équitable. C'est pourquoi tous les travaux de candidats doivent être disponibles jusqu'à la publication des résultats.

Cette situation a parfois lieu parce que plusieurs enseignants ont noté les travaux et qu'ils n'ont pas réussi à établir une norme commune. Étant donné qu'un facteur de révision de notation unique est appliqué à chaque composante d'évaluation interne dans l'établissement, il est fondamental que les enseignants harmonisent leur notation avant d'envoyer les travaux à l'IB.

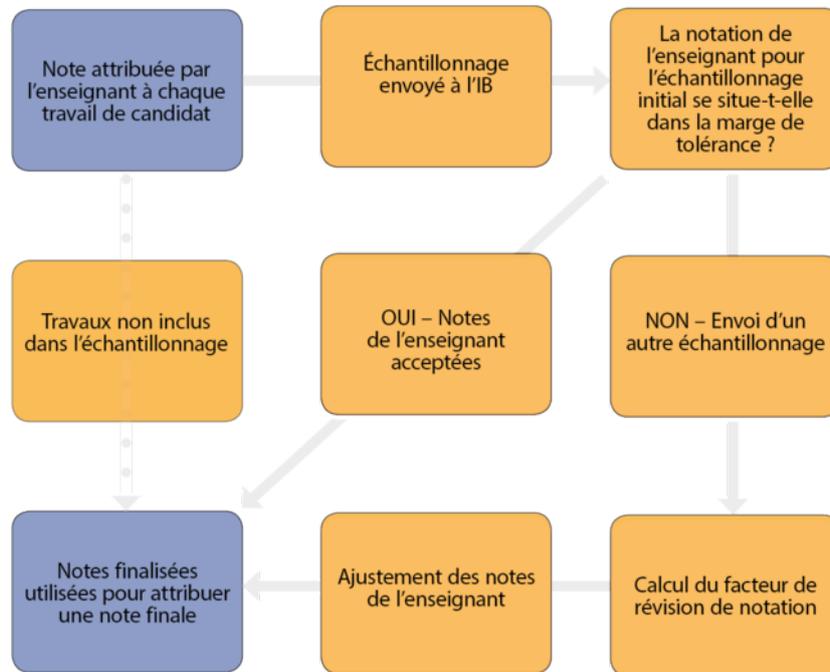
Échantillonnage dynamique

- La révision de notation reposant sur l'échantillonnage dynamique signifie que si la notation d'un enseignant se situe dans la marge de tolérance, ses notes sont acceptées.
- Si la notation d'un enseignant se situe en dehors de la marge de tolérance, un facteur de révision de notation arithmétique est appliqué.
- Chaque examinateur vérifiant les notes d'évaluation interne attribuées par un enseignant est soumis à un contrôle à l'aide d'un modèle d'assurance de la qualité similaire à celui utilisé pour les examens.

Le principe de la révision de notation reposant sur l'échantillonnage dynamique est le même que celui suivi pour les examinateurs. Ainsi, si un enseignant montre qu'il a compris la norme globale en restant dans la marge de tolérance dans tout l'échantillonnage, il est reconnu que toutes ses notes sont conformes à la norme exigée. Si ses notes sont en dehors de la marge de tolérance, un facteur de révision de notation est calculé et appliqué. Ce système est expliqué dans le schéma qui suit.

Figure 51

Schéma de la révision de notation reposant sur l'échantillonnage dynamique



À la différence de l'approche précédente, avec la révision de notation reposant sur l'échantillonnage dynamique, il est fort peu probable que les établissements voient un petit facteur de révision de notation être appliqué à leur notation, car, dans ce cas, il sera considéré que la notation se situe dans la marge de tolérance. Une autre différence est que l'IB sera en mesure de faire un retour d'information bien plus détaillé, étant donné que l'analyse d'un seul examinateur sera requise du fait du modèle d'assurance de la qualité présenté ci-après.

Pour que la révision de notation soit juste, les examinateurs doivent comprendre la norme globale définie par l'examineur principal et la respecter pour procéder à la révision. Pour veiller à ce que cela se produise, l'IB suit un système similaire au modèle d'assurance de la qualité du travail des examinateurs.

Les examinateurs qui participeront à la révision de notation suivent une formation sur la norme globale, puis il est vérifié qu'ils procèdent à la révision de notation en respectant cette norme. Comme il est décrit dans la section « [Le modèle d'assurance de la qualité](#) », l'examineur principal prépare trois types de copies révisées de manière définitive :

- des copies d'entraînement, pour expliquer la norme globale ;
- des copies de validation, pour vérifier que les examinateurs comprennent la norme globale ;
- des copies de contrôle, pour vérifier que les examinateurs respectent systématiquement la norme globale.

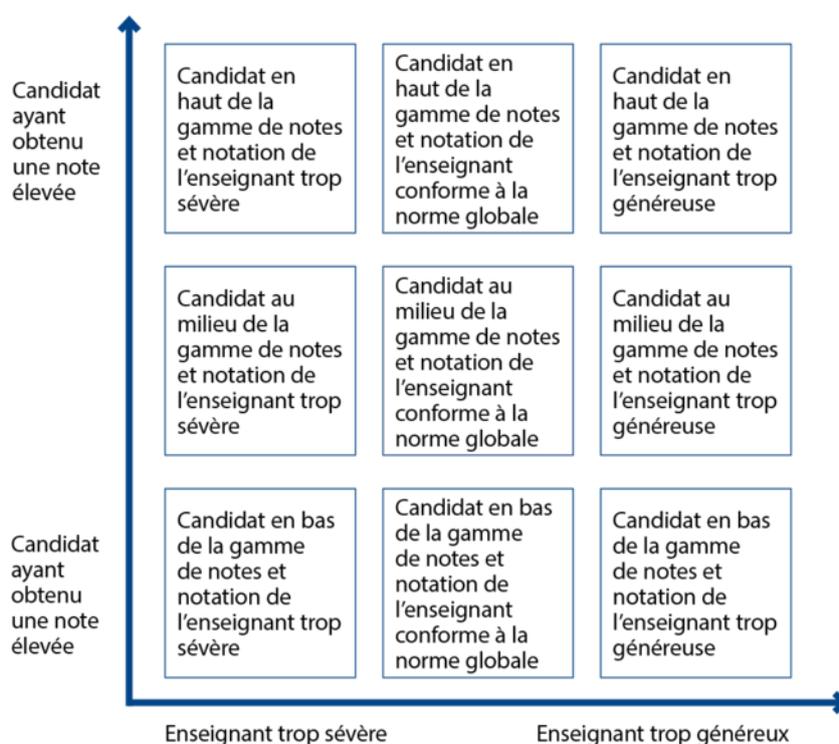
Contrairement à ce qui se produisait avec la précédente approche de la révision de notation, avec l'échantillonnage dynamique, les examinateurs reçoivent des travaux issus d'établissements très divers sans qu'aucun ordre précis ne soit suivi. Cela réduit les risques qu'un examinateur se forge une opinion sur la notation de l'enseignant (qu'il la juge trop généreuse ou trop sévère) dès le premier candidat et qu'il recherche ensuite cette tendance dans le reste de l'échantillonnage de l'établissement. Par ailleurs, cette méthode permet d'inclure des copies de contrôle sans qu'elles apparaissent comme telles de façon évidente.

Le même examinateur continuera à réviser tous les travaux d'un établissement, mais il ne les révisera plus tous ensemble, et il devra fournir des commentaires sommaires sur la notation de l'établissement une fois qu'il aura été décidé s'il est nécessaire d'appliquer un facteur de révision de notation.

Le modèle d'assurance de la qualité est plus complexe avec l'échantillonnage dynamique qu'avec l'autre méthode de notation. En effet, avec les épreuves d'examen, l'examinateur doit situer un travail par rapport à toute une gamme de notes, alors qu'aux fins de la révision de notation, l'IB doit également s'assurer que l'examinateur est capable de déterminer si les notes de l'enseignant sont trop sévères, trop généreuses ou correctes. Le diagramme ci-dessous illustre ces nuances.

Figure 52

Éventail requis de travaux révisés de manière définitive

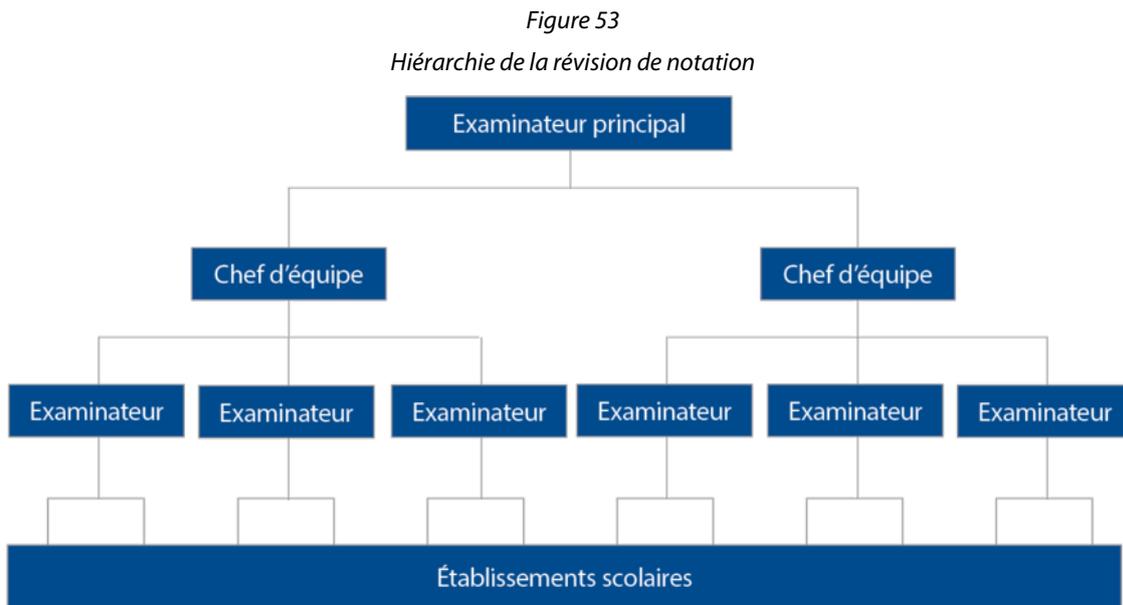


La hiérarchie de la révision de notation selon l'ancien système

- Le système de hiérarchie de la révision de notation est en train d'être remplacé par l'échantillonnage dynamique, qui permet à l'IB de s'assurer que tous les examinateurs respectent la norme globale avant de procéder à la révision de notation.
- Avec l'ancien système, chaque examinateur se trouvant en dessous de l'examinateur principal voyait également un facteur de révision de notation être appliqué à ses notes. Le facteur de révision de notation final appliqué aux notes de l'enseignant était une combinaison de tous ces facteurs. Ainsi, chaque série de notes d'un enseignant était alignée sur la norme globale établie par l'examinateur principal.

Avant d'introduire l'échantillonnage dynamique, l'IB se servait d'une deuxième approche de la révision de notation pour vérifier que les examinateurs révisaient correctement la notation des enseignants. L'examineur principal révisait un échantillonnage de notes attribuées par son équipe d'examineurs superviseurs, et un ajustement était apporté à ces notes pour garantir leur adéquation avec la norme de l'examineur principal.

Cela donnait lieu à un processus hiérarchique garantissant que les notes définitives attribuées à chaque établissement étaient conformes à la norme de notation fixée par l'examineur principal. Les différents niveaux de la hiérarchie sont illustrés ci-après dans le cas d'une composante évaluée en interne à laquelle de nombreux candidats sont inscrits.



Le schéma ci-dessus illustre les points suivants.

- Les notes d'un établissement peuvent être ajustées sur la base de l'échantillonnage envoyé à l'examineur.
- La notation de chaque examinateur est également révisée et ajustée sur la base d'un échantillonnage de sa notation, envoyé à un examinateur superviseur (« chef d'équipe »).
- Les notes du chef d'équipe peuvent à leur tour être ajustées sur la base d'un échantillonnage de sa notation envoyé à l'examineur principal.

Par conséquent, il existe une chaîne de révision de notation, qui permet d'appliquer une série d'ajustements aux notes d'un établissement avant que ne soient attribuées les notes définitives révisées (alignées sur la norme de l'examineur principal).

Les commentaires sur l'évaluation interne

La révision de notation de l'évaluation interne vise à garantir que tous les enseignants suivent la même norme de notation, et l'IB aimerait pouvoir éviter d'avoir à appliquer des facteurs de révision de notation aux travaux d'un établissement. Pour que les enseignants puissent comprendre où ils se situent par rapport à la norme globale, l'IB fournit un retour d'information sur la notation des tâches d'évaluation interne. Cela permet aux enseignants de comprendre les raisons de l'application d'un facteur de révision de notation.

Le retour d'information qui est fait aux établissements ne vise pas à expliquer comment les candidats de l'échantillonnage auraient pu obtenir de meilleurs résultats.

L'attribution des notes finales (et la totalisation)

- En quoi consiste l'attribution des notes finales ?
- Les notes finales doivent avoir la même signification, quelle que soit la session pendant laquelle un candidat passe ses examens.
- Le processus d'attribution des notes finales détermine comment convertir les notes en notes finales pour garantir que ce soit le cas.
- Les seuils d'attribution des notes finales sont fixés à partir d'un éventail de preuves, notamment l'avis de spécialistes sur les travaux des candidats et les résultats de la cohorte.
- La note finale globale attribuée à un candidat est celle qui revêt la plus grande importance, et non la note accordée à chaque composante.

Il convient de garder à l'esprit que notes et notes finales sont deux choses différentes. Veuillez vous référer à la section « [Notes et notes finales sont deux choses différentes](#) » pour comprendre les raisons de cette distinction. Le processus d'attribution des notes finales est un mécanisme permettant de décider comment convertir les notes en notes finales. Cette décision est prise à l'issue de plusieurs jours de discussions entre l'examineur principal, l'examineur en chef, les examinateurs superviseurs et l'IB. Diverses preuves (décrites plus en détail dans la section « [Les preuves utilisées pour l'attribution des notes finales](#) ») sont étudiées pour convenir de ce qui est juste pour les candidats de l'année en cours, mais aussi pour ceux qui ont été évalués dans cette matière les années précédentes. Enfin, l'examineur en chef fait part à l'IB des résultats qu'il recommande pour la session d'examens concernée.

Le processus d'attribution des notes finales permet également de recueillir des preuves de l'efficacité des évaluations. Ces informations sont ensuite utilisées pour améliorer les sessions suivantes. Plus précisément, la première année d'évaluation d'un nouveau cours, le responsable des programmes d'études de l'IB, qui est chargé du développement du programme d'études du cours, apportera son soutien lors de la réunion de délibérations afin de faciliter l'alignement entre les résultats de l'évaluation et l'intention nourrie pour les objectifs globaux du cours, et d'orienter le prochain développement des programmes.

L'objet principal de ces discussions est l'établissement des « seuils d'attribution des notes finales », autrement dit la définition du nombre de points minimum requis pour pouvoir attribuer chaque note finale aux candidats. Parce qu'il n'est pas pratique d'émettre un jugement détaillé pour chaque note finale dans chaque matière, l'IB demande aux examinateurs de recommander plusieurs « seuils d'attribution des notes finales fondés sur une appréciation » et calcule lui-même les autres seuils de façon arithmétique. Dans le PEI, le Programme du diplôme et le POP, pour lesquels il existe sept notes finales, ces seuils d'attribution des notes finales fondés sur une appréciation sont les suivants : 2 – 3, 3 – 4 et 6 – 7. Ces seuils représentent en effet le plus haut niveau de compréhension et de connaissance d'une matière (6 – 7), les résultats d'un candidat possédant une connaissance et une compréhension élémentaires (2 – 3), et le niveau d'une compréhension et d'une connaissance assurées (3 – 4).

Les seuils d'attribution des notes finales doivent être modifiés chaque session. Cela s'explique par le fait que les questions posées aux candidats seront elles aussi modifiées et que les seuils d'attribution des notes finales devront être adaptés pour tenir compte des tâches plus faciles ou plus difficiles. L'IB fait son possible pour garantir que ses épreuves d'examen présentent le même niveau de difficulté chaque année. Néanmoins, en raison des enjeux élevés, il ne peut procéder à aucune mise à l'essai des examens avant la session, par crainte d'une divulgation des questions.

Lorsqu'ils décident des notes finales à attribuer, les examinateurs et l'IB doivent prendre en considération les aspects qui suivent.

Figure 54

Aspects à prendre en considération pendant le processus d'attribution des notes finales

Cohorte	• En quoi la cohorte qui passe les examens peut-elle être comparée années aux cohortes des années précédentes ?
Examen	• Dans quelle mesure l'examen a-t-il rempli les attentes de l'auteur ? • Qu'ont pensé les enseignants du niveau de difficulté de l'examen ?
Copies d'examen	• Examen des travaux des candidats et comparaison avec ceux des années précédentes et les descripteurs des notes finales.
Résultats	• En quoi les résultats des candidats sont-ils comparables à ceux des années précédentes ?
Équilibre	• Toutes les preuves disponibles conduisent-elles à la même conclusion ou sont-elles contradictoires ?
Cas spéciaux	• Examen des cas des candidats ayant produit des réponses inhabituelles ou des candidats affectés par des circonstances exceptionnelles.

Bien souvent, l'IB doit combiner plusieurs composantes d'évaluation pour parvenir à une note finale globale. Le détail de cette manœuvre est expliqué dans la section « La totalisation » ci-après. Mais l'élément fondamental, que les examinateurs doivent garder à l'esprit pendant la réunion d'attribution des notes finales, est que le résultat global importe le plus. Si nécessaire, certaines composantes peuvent être moins parfaites pour atteindre un résultat global juste.

Officiellement, le processus d'attribution des notes finales remplit plusieurs objectifs.

- Il permet de déterminer, à partir des travaux des candidats, ce qui indique un changement dans la qualité du travail justifiant de passer à un autre descripteur des notes finales.
- Pour chaque composante, il convient de déterminer les notes qui correspondront aux différents seuils d'attribution des notes finales.
- Il faut garantir que l'ensemble des seuils établis pour chaque cours représente une juste attribution des notes finales.
- Il permet de décider de la note finale à attribuer aux réponses des candidats considérées comme atypiques ou inhabituelles.

Le processus d'attribution des notes finales sera réussi si :

- les jugements portés sur les copies et les preuves tirées des résultats concordent largement, en tenant compte des informations sur la cohorte ;
- les écarts concernant les résultats des établissements peuvent être expliqués ;
- l'examineur en chef et le chef de la direction de l'évaluation de l'IB sont certains que les normes d'évaluation ont été respectées.

Les seuils d'attribution des notes finales fondés sur une appréciation et calculés par interpolation

Les seuils d'attribution des notes finales fondés sur une appréciation sont ceux qui sont recommandés par l'examineur en chef, à partir des discussions tenues lors du processus d'attribution des notes finales. Dans le Programme du diplôme, le POP et le PEI, ces seuils d'attribution des notes finales fondés sur une appréciation sont les suivants : 2 – 3, 3 – 4 et 6 – 7. Ces seuils représentent en effet le plus haut niveau de compréhension et de connaissance d'une matière (6 – 7), les résultats d'un candidat possédant une connaissance et une compréhension élémentaires (2 – 3), et le niveau d'une compréhension et d'une connaissance assurées (3 – 4).

Les autres seuils, à savoir 1 – 2, 4 – 5 et 5 – 6, sont calculés de façon arithmétique à partir des seuils d'attribution des notes finales fondés sur une appréciation ; c'est pourquoi ils sont qualifiés de « seuils d'attribution des notes finales calculés par interpolation ».

Même si les seuils calculés par interpolation reposent sur des preuves issues de l'analyse des travaux des candidats, ils jouent un rôle important dans l'examen de l'accomplissement global des candidats. En effet, en cas de variation importante du nombre de candidats ayant obtenu une note finale de 5 ou de 1, une discussion doit être menée en tenant compte de la cohorte et peut conduire à une révision des seuils fondés sur une appréciation ou, dans des circonstances exceptionnelles, à un réexamen des travaux des candidats se situant à ces seuils.

L'impact de l'évaluation électronique sur l'attribution des notes finales

Le processus d'attribution des notes finales doit avoir lieu, et ce, quelle que soit la forme de l'évaluation. L'introduction de l'évaluation électronique ne modifiera donc en rien les principes d'attribution des notes finales.

L'évaluation électronique pourrait avoir un impact sur le processus de notation en ce sens qu'il permettra de proposer des évaluations mesurant plus efficacement l'atteinte des objectifs du cours. Il est souvent difficile pour les examinateurs d'observer des preuves de l'analyse, de la recherche ou de la réflexion collective dans les épreuves sur papier, ce qui peut impliquer une sous-représentation de ces aspects des descripteurs des notes finales. L'évaluation électronique peut aider à résoudre ce problème.

Les preuves utilisées pour l'attribution des notes finales

L'attribution des notes finales est un processus reposant sur des preuves, qui doit exploiter diverses informations, et notamment :

- les commentaires des enseignants sur l'évaluation ;
- les jugements spécialisés des examinateurs sur les exemples de travaux de candidats ;
- l'analyse des informations statistiques comparant les résultats des candidats de l'année en cours à ceux des candidats des années précédentes.

Aucun type de preuve n'a plus d'importance que les autres : toutes doivent peser de façon égale pour tirer des conclusions.

L'examen de la cohorte de l'année en cours

La première étape du processus d'attribution des notes finales consiste à se demander en quoi la cohorte qui passe l'évaluation se rapproche des cohortes des années précédentes. Si la plupart des établissements ayant inscrit des candidats pour l'année en cours sont les mêmes que les années précédentes, il est probable que les différences de réussite s'expliqueront par la difficulté des épreuves d'examen. Par ailleurs, si un grand nombre de nouveaux établissements inscrivent des candidats à une matière pour la première fois, ou si un pourcentage élevé de candidats repassent l'évaluation, l'IB pourrait décider qu'il est nécessaire que les notes finales tiennent compte des différences de réussite.

Parmi les facteurs qui seront pris en considération par les examinateurs lors de la comparaison des cohortes passant l'évaluation, il convient de citer :

- les variations du nombre de candidats passant les évaluations ;
- les variations du pourcentage de candidats passant les évaluations en anglais, en espagnol et en français (et dans d'autres langues, le cas échéant) ;
- le nombre de nouveaux établissements et le nombre d'élèves dans ces nouveaux établissements ;

- tout changement concernant les options choisies par les élèves.

Le retour d'information sur l'évaluation

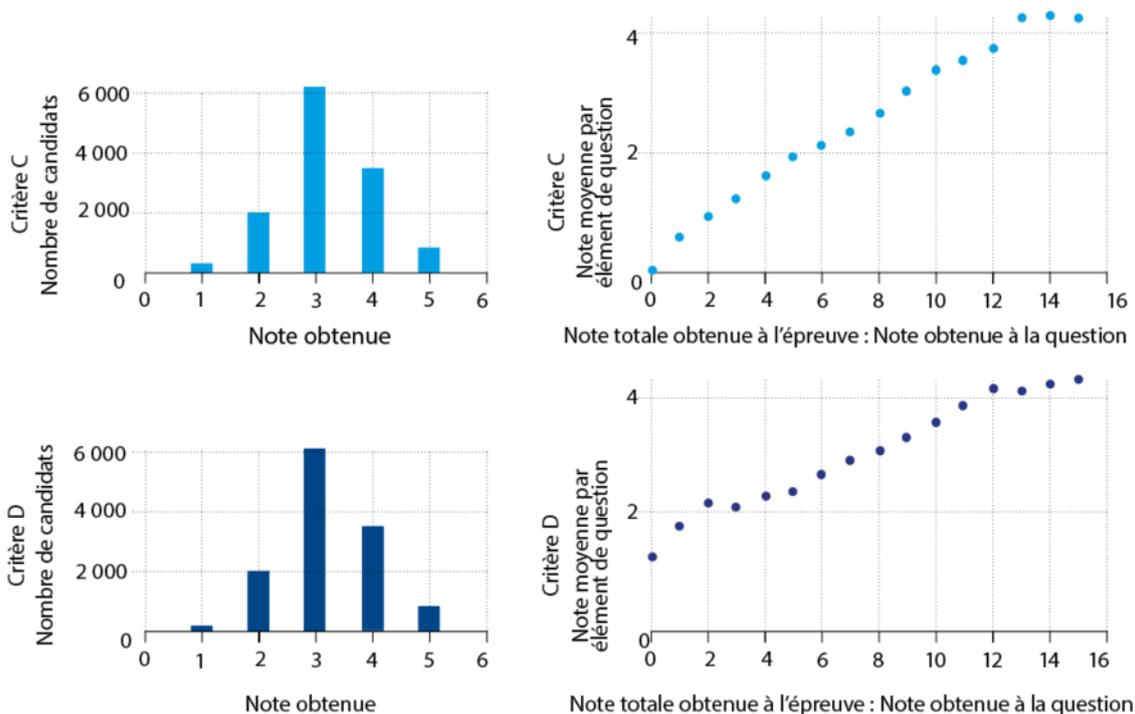
L'étape suivante consistera à se demander comment s'est déroulée l'évaluation par rapport aux attentes. Si une question en particulier se révèle bien plus difficile que prévu, ou si des éléments de question censés faire une distinction entre les candidats méritant la note finale 6 et ceux méritant la note finale 7 n'y parviennent pas, les examinateurs superviseurs devront en tenir compte lors de l'établissement des seuils d'attribution des notes finales.

Les examinateurs superviseurs utiliseront les commentaires des enseignants sur les épreuves d'examen (essentiellement les formulaires G2) comme base de leur discussion ainsi que les statistiques sur les résultats obtenus à chaque question ou élément (voir le diagramme ci-après). Ils utiliseront également les résumés des rapports provenant des examinateurs.

Figure 55

Exemple de statistiques à la disposition des examinateurs concernant les éléments

Analyse statistique des éléments							Corrélation entre l'élément et le reste de la composante	
Question	Nombre de réponses	Nombre total de candidats à l'examen	Pourcentage de candidats ayant tenté de répondre à la question	Note moyenne par élément de question	Note la plus haute obtenue par élément de question	Écart type de la population des notes obtenues aux éléments de question	Question	Coefficient de corrélation
01 Critère A	12 071	42 042	28,71 %	2,7	5	0,95	01 Critère A	0,8044
01 Critère B	12 071	42 042	28,71 %	2,6	5	0,94	01 Critère B	0,8018
01 Critère C	12 071	42 042	28,71 %	2,8	5	0,84	01 Critère C	0,7983
01 Critère D	12 071	42 042	28,71 %	3,2	5	0,79	01 Critère D	0,7253
02 Critère A	29 848	42 042	71,00 %	2,8	5	0,98	02 Critère A	0,7954
02 Critère B	29 848	42 042	71,00 %	2,7	5	0,93	02 Critère B	0,7944
02 Critère C	29 848	42 042	71,00 %	2,9	5	0,83	02 Critère C	0,7865
02 Critère D	29 848	42 042	71,00 %	3,2	5	0,79	02 Critère D	0,7320



Les examinateurs superviseurs s'appuieront également sur leur expérience de la notation des travaux des candidats lors de la session. Il est donc indispensable que les discussions sur l'attribution des notes finales réunissent des personnes ayant noté chacune des composantes au cours de la session. Même s'il préside les discussions, l'examinateur en chef ne doit pas nécessairement avoir corrigé des travaux de candidats. Il peut en effet être utile de ne pas avoir de préjugés. Cela impose toutefois une responsabilité supplémentaire aux examinateurs principaux, qui doivent impérativement apporter tous les renseignements utiles sur la qualité générale des travaux des candidats.

La recherche de preuves dans les copies

Les descripteurs des notes finales expliquent les caractéristiques qui doivent se retrouver dans les travaux des candidats pour chacune des notes finales. Lors de cette étape du processus d'attribution des notes finales, il est essentiel que les examinateurs se concentrent non pas sur les notes attribuées, mais sur la nature des réponses des candidats et la mesure dans laquelle elles correspondent aux descripteurs des notes finales. Par conséquent, on sélectionne autant que possible des copies d'examen dans lesquelles les candidats ont fourni des réponses équilibrées tout au long de l'épreuve plutôt que des copies dans lesquelles les candidats ont obtenu de bonnes notes pour certaines questions et de mauvaises notes pour d'autres.

Avant de commencer à passer en revue les copies, il est souvent utile de revenir aux exemples de travaux des années précédentes, afin de rappeler aux examinateurs les attentes correspondant aux différentes notes finales. Les copies utilisées pour déterminer les seuils d'attribution des notes finales doivent être disponibles pendant tout le processus d'attribution des notes finales.

Avant la réunion de délibérations, l'équipe d'examinateurs superviseurs, et surtout les examinateurs principaux, propose des seuils d'attribution provisoires des notes finales pour chaque composante, en indiquant au niveau de quelles notes ils estiment nécessaire de fixer les seuils, sur la base de leurs expériences passées des niveaux attendus. Après avoir pris connaissance de ces seuils provisoires, de l'avis partagé par les participants sur la façon dont chaque épreuve s'est déroulée et de la distribution globale des notes, le responsable de matière de l'IB peut suggérer la gamme de notes (copies de l'échantillonnage)

dont les examinateurs superviseurs auront besoin pour commencer leur examen des travaux des candidats pour chaque note finale.

Chaque examinateur superviseur doit passer en revue les copies de candidats sélectionnées et déterminer quel descripteur des notes finales reflète le mieux la qualité du travail. Il est important qu'il se concentre sur la copie dans son intégralité et qu'il ne se laisse pas influencer par les notes attribuées. Il est utile de savoir quelles questions ont été conçues pour apporter des preuves de la présence des qualités associées à la note finale la plus haute, mais ces questions doivent simplement aider à prendre une décision globale. Cette tâche s'avère compliquée, car même les candidats qui présentent des réponses d'un niveau relativement homogène montreront souvent des qualités associées à diverses notes finales au fil de leurs réponses. Les examinateurs devront donc juger quelle est la note finale qui leur correspond le mieux. Il est acceptable, et même souvent utile, d'indiquer si une copie correspond tout juste à un descripteur des notes finales ou si elle est proche de la note finale supérieure suivante (ce qui est souvent indiqué sous la forme 7- ou 6+ par exemple).

Il est essentiel que les examinateurs évitent tout type de préjugé ou de biais pour mener cette tâche à bien ; c'est pourquoi ils ne doivent pas partager leurs points de vue avant que chacun ait consigné son résultat de son côté. De même, il est utile de ne pas informer les examinateurs (à l'exception de l'examinateur en chef) des conclusions de l'analyse statistique avant la fin de cette étape.

Déterminer la note finale qui correspond le mieux à une copie est un exercice très subjectif, d'où la grande possibilité d'écart entre les examinateurs. Les notes et les notes finales n'illustrent pas tout à fait le même accomplissement. Il n'est donc pas aberrant d'attribuer à un candidat, qui a montré une bonne compréhension mais fait quelques erreurs, une note finale plus élevée qu'à un candidat qui a excellé pour la tâche la plus facile (et obtenu de ce fait beaucoup de points), mais n'a pas montré le même niveau de compréhension, même si celui-ci a obtenu un peu plus de points. Il ne faut pas non plus perdre de vue que cet examen des copies ne concerne qu'un petit nombre d'exemples de travaux de candidats et que toutes les décisions doivent en tenir compte.

Des études ont révélé que les examinateurs étaient très doués pour repérer les copies affichant un niveau de qualité différent des autres copies déjà examinées (voir la section « [Les autres formes de notation](#) »). L'IB demande donc aux examinateurs de commencer avec la note la plus élevée de l'échantillonnage et de passer en revue les copies jusqu'à ce qu'ils estiment ne plus trouver constamment des preuves des qualités associées à la note finale la plus élevée. Ils doivent ensuite commencer par la note la plus basse de l'échantillonnage et passer en revue les copies jusqu'à ce qu'ils trouvent constamment des preuves des qualités associées à la note finale la plus élevée. Si nécessaire, la gamme de notes de l'échantillonnage peut être augmentée.

Lorsque toutes les décisions des examinateurs superviseurs en matière d'attribution d'une note finale sont rassemblées, elles doivent faire apparaître la gamme de notes au niveau de laquelle situer un seuil d'attribution des notes finales. Il s'agit d'une « zone d'incertitude », qui va de la note la plus basse pour laquelle suffisamment de preuves existent pour situer le seuil d'attribution des notes finales à la note la plus élevée. Aucune définition précise n'indique comment décider de la zone d'incertitude. Les examinateurs superviseurs doivent en discuter ensemble, en s'appuyant sur les décisions prises par chacun.

Figure 56

Exemple de résultats de l'analyse des copies d'examen (Lors de la réunion d'attribution des notes finales, un nombre plus élevé de copies d'examen est utilisé pour chaque note.)

Numéro de la copie d'examen	Note	Examinateur 1 (examinateur principal)	Examinateur 2	Examinateur 3	Examinateur 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 d'incertitude

Seuls les questionnaires à choix multiple ne sont pas visés par ce processus. L'expérience montre en effet qu'il est très difficile d'émettre des jugements sur les seuils d'attribution des notes finales à partir de la qualité du travail des candidats lorsque les épreuves sont composées uniquement de questions à choix multiple. Cela peut s'expliquer par le fait que les réponses contiennent peu de preuves des réalisations réelles des candidats, sur lesquelles baser un jugement. Pour ce type d'épreuve, les seuils d'attribution des notes finales sont calculés de manière à ce que le pourcentage de candidats à chaque note finale soit le plus proche possible de celui obtenu à partir d'une appréciation pour l'épreuve d'examen qui s'en rapproche le plus.

L'analyse des statistiques sur les résultats

On pourrait croire que dans un système d'évaluation critériée dépendant du jugement professionnel des correcteurs, les examinateurs superviseurs devraient être capables d'établir des seuils d'attribution des notes finales en considérant simplement les questions de l'épreuve d'examen et les réponses attendues pour chaque question. En réalité, il est très difficile d'établir ces seuils avec précision avant de connaître la façon dont les candidats ont répondu. Cresswell (2000) est parvenu à la conclusion que les correcteurs finaux parvenaient le plus souvent à déterminer si des épreuves étaient plus faciles ou plus difficiles d'une session d'examens à une autre, mais pas à estimer dans quelle mesure elles étaient plus faciles ou plus difficiles.

[...] il existe de bonnes raisons théoriques et empiriques de croire que « le maintien des normes d'après la définition de l'approche critériée limitée » est trop difficile pour les correcteurs finaux, même expérimentés. De plus, des preuves empiriques indiquent que les jugements en matière d'évaluation des correcteurs finaux sont influencés par des facteurs auxquels ils ne devraient pas être sensibles, comme la cohérence des notes dans une copie. Cela constitue de bonnes raisons de douter du fait que les notes finales décidées par les correcteurs finaux suffiront, à elles seules, à soutenir les jugements portés sur les examens.

(Baird, Cresswell et Newton, 2000, p. 213 – 229)

L'objectif de la réunion d'attribution des notes finales étant de maintenir une signification ou un niveau cohérents des notes finales, il est intéressant de réfléchir à l'une des définitions de la comparabilité données par Cresswell.

Deux examens peuvent être considérés comme de niveau comparable si deux groupes de candidats affichant le même éventail d'aptitudes et les mêmes résultats antérieurs, fréquentant des établissements semblables qui appliquent les mêmes politiques d'admission, étudiant avec des enseignants aux compétences égales, et présentant le même niveau de motivation, reçoivent des notes finales qui suivent strictement la même répartition, après avoir étudié leurs programmes respectifs et passé les examens.

(Cresswell, 1996, p. 57 – 84)

Pour faciliter leur travail d'attribution des notes finales, des « seuils recommandés statistiquement » sont fournis aux examinateurs superviseurs. Il s'agit des seuils qui produiraient exactement le même pourcentage cumulé que l'année précédente pour la note finale concernée. Le terme « pourcentage cumulé » désigne le nombre total de candidats ayant obtenu une note finale précise ou celles au-dessus. Il est utilisé pour expliquer le fait que, si les candidats les plus compétents obtiennent de meilleurs résultats (ce qui fait augmenter le nombre de candidats avec une note finale de 7), il est possible de s'attendre à ce qu'il y ait moins de candidats avec une note finale de 6.

Les examinateurs participant au processus d'attribution des notes finales reçoivent aussi des informations sur la note moyenne obtenue par les candidats et des histogrammes illustrant la répartition réelle des notes. Même si ces informations sont prises en compte dans les seuils recommandés statistiquement, il est souvent utile de les analyser de près.

Pour revenir aux deux citations précédentes, il est important de garder à l'esprit qu'un groupe de candidats différent passera les examens chaque session. De ce fait, il est probable que les postulats de la seconde citation ne soient pas totalement vérifiés. Cela sera notamment le cas si les deux cohortes sont très différentes. De même, lorsqu'il existe un très grand nombre de candidats avec des expériences scolaires similaires (par exemple, scolarisation dans des établissements autorisés à proposer un programme de l'IB), il est plus probable que les différences de résultats notables soient dues à la décision de répondre à des séries de questions différentes plutôt qu'aux écarts entre les accomplissements.

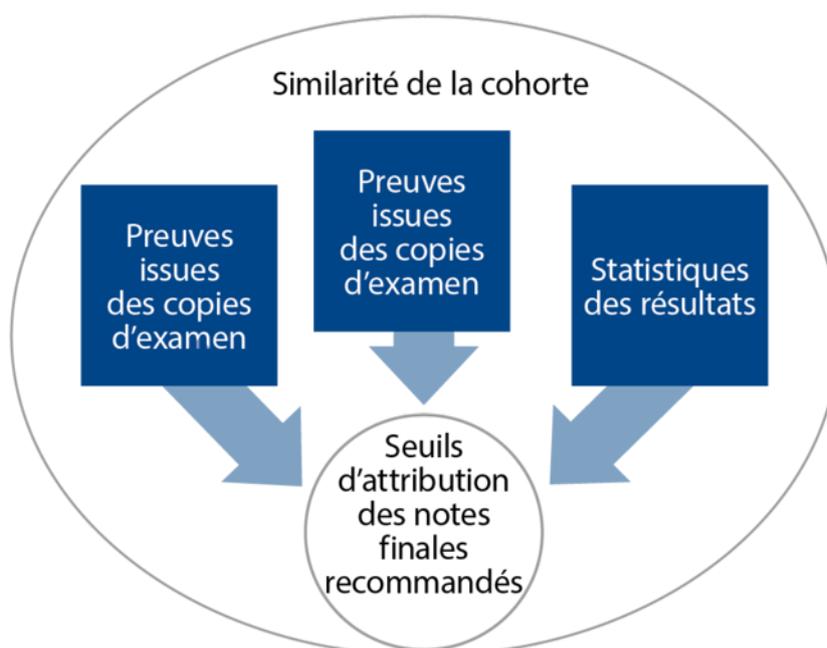
La taille de la cohorte doit être prise en considération lors de l'examen de l'importance des seuils recommandés statistiquement, et l'IB ne suit aucune règle formelle sur la question, car d'autres facteurs sont aussi importants. En règle générale, si les candidats se comptent en dizaines seulement, des écarts considérables pourront être observés au niveau du résultat global, mais il est assez peu probable que cela se produise s'ils sont des milliers.

Trouver un juste équilibre entre toutes les preuves

Aucun type de preuve n'a plus d'importance que les autres. Le travail de l'examineur en chef et de son équipe consiste à trouver un juste équilibre entre toutes les preuves pour faire une recommandation.

Figure 57

Preuves venant étayer le choix des seuils d'attribution des notes finales



Souvent, les différentes preuves suggèrent les mêmes résultats (par exemple, les seuils recommandés statistiquement se situent dans la zone d'incertitude), mais il peut arriver que certains soient en contradiction. Dans de tels cas, il est essentiel d'expliquer cet écart dans la justification de la décision finale.

Lors de cette dernière étape du processus d'attribution des notes finales, il est possible de montrer les effets éventuels des différentes décisions sur les résultats globaux de la cohorte. Conformément au principe selon lequel le résultat global importe plus que celui obtenu dans les différentes composantes, l'examineur en chef pourrait comparer la note finale moyenne de la cohorte aux prévisions des enseignants (en se référant aux années précédentes pour comprendre la fiabilité d'une telle comparaison). Une autre méthode consisterait à exclure les nouveaux établissements des résultats globaux pour vérifier s'ils affichent un niveau très différent des établissements plus expérimentés.

Enfin, l'examineur en chef doit soumettre ses recommandations, accompagnées des justifications nécessaires, au chef de la direction de l'évaluation de l'IB.

Les descripteurs des notes finales

Les descripteurs des notes finales sont une synthèse des caractéristiques du niveau d'accomplissement attendu pour chaque note finale. Ils occupent une place fondamentale dans l'établissement des seuils d'attribution des notes finales, étant donné qu'ils décrivent ce que les examinateurs doivent chercher quand ils passent en revue le travail des candidats. Il faut donc qu'ils soient rédigés pour chaque matière, afin de garantir qu'ils seront appliqués de la même manière d'une session d'examens à l'autre.

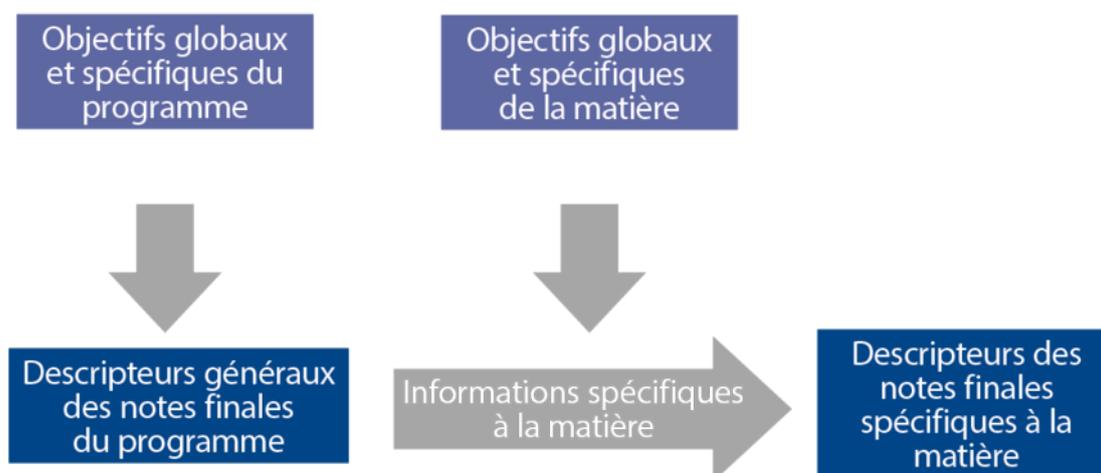
L'importance du rôle des descripteurs des notes finales est telle qu'elle fait d'eux une pièce maîtresse de la validité de l'évaluation des cours de l'IB. Ces descripteurs doivent refléter les objectifs globaux et spécifiques de la matière, afin de garantir que l'IB évalue et récompense ce qui doit l'être.

Ils jouent également un rôle essentiel en favorisant la comparabilité entre les matières, car ils servent de référence commune pour toutes les matières. Il est donc nécessaire d'avoir des descripteurs généraux des notes finales de façon à ce que la note finale 4, par exemple, ait la même signification dans toutes les matières.

Ces principes impliquent d'avoir un lien évident entre les objectifs du programme de l'IB et les descripteurs généraux des notes finales ainsi qu'une évolution logique vers les descripteurs des notes finales spécifiques à la matière. Le diagramme ci-dessous illustre cette idée.

Figure 58

Lien entre les descripteurs généraux des notes finales et les descripteurs des notes finales spécifiques à la matière



Les seuils d'attribution des notes finales fixes

- Lorsque les mêmes tâches sont utilisées chaque session, il est attendu que les seuils d'attribution des notes finales restent les mêmes.
- S'il est démontré que les normes sont incorrectes ou qu'elles ont changé au fil du temps, l'IB reverra et modifiera ces seuils « fixes ».

Pour l'essentiel de l'évaluation interne, et une partie des composantes notées en externe, il est demandé aux candidats d'effectuer des tâches presque identiques chaque session. C'est notamment le cas avec la préparation d'un portfolio d'arts et le projet personnel, pour ne citer que ces exemples. Dans de telles situations, il peut raisonnablement être supposé, puisque l'IB maintient la même norme de notation chaque année (grâce à la normalisation à l'aide des travaux des années précédentes), que les seuils d'attribution des notes finales resteront eux aussi les mêmes.

Lors du processus d'attribution des notes finales, l'examineur principal et l'examineur en chef devront se demander si des preuves indiquent que les seuils d'attribution des notes finales existants ne sont pas appropriés. Dans la plupart des cas, ils devraient conclure que les seuils existants peuvent être conservés.

Malgré cela, il ne doit faire aucun doute que ces seuils sont vérifiés chaque année et qu'ils peuvent être modifiés si des éléments prouvent qu'ils ne sont plus efficaces pour maintenir la comparabilité des notes finales entre les sessions. Cela pourrait se produire en cas de changement important du comportement des candidats (par exemple, l'influence des progrès technologiques sur la réalisation d'une tâche), de changements dans les travaux remis qui renforcent l'alignement sur le barème de notation sans améliorer la qualité globale du travail des candidats, ou de révision d'une approche de l'IB en matière de notation ou de révision de notation des travaux qui modifie la norme de notation.

La totalisation

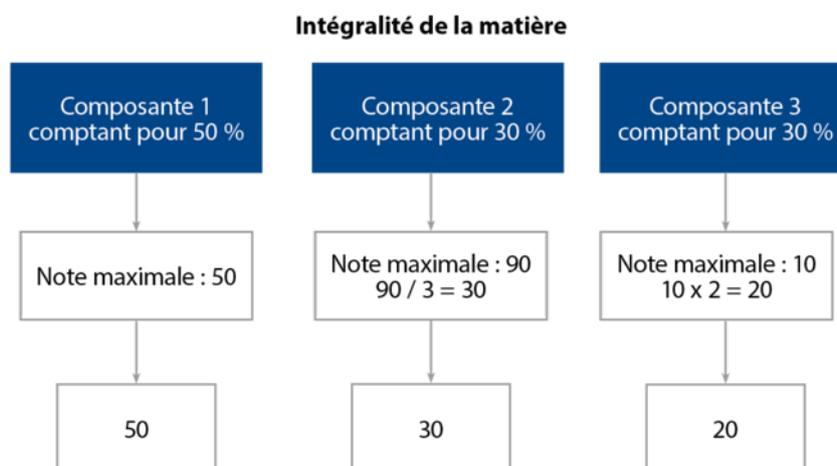
- Le terme « totalisation » désigne le processus d'addition des notes obtenues dans les composantes pour produire un résultat général.
- Pour que chaque composante contribue à la note définitive dans les bonnes proportions (pondération), il peut s'avérer nécessaire d'appliquer un coefficient aux notes des composantes.
- L'IB utilise un « système de compensation » grâce auquel les candidats peuvent compenser de mauvais résultats obtenus dans une question ou une composante par un niveau plus élevé atteint dans une autre.
- La note finale par matière d'un candidat est calculée en totalisant les notes obtenues dans chaque composante, et non ses notes finales par composante.

Le terme « totalisation » désigne le processus d'addition des notes (et des seuils) issues des différentes composantes pour produire une note définitive ou un seuil d'attribution des notes finales global. Pour y parvenir, il peut s'avérer nécessaire d'appliquer un coefficient aux notes globales (ou seuils) des composantes.

Un coefficient est appliqué afin de conserver la pondération voulue pour chaque composante pour ce qui est de sa contribution à l'évaluation globale du cours. Cela implique de multiplier ou de diviser les notes des composantes pour qu'elles contribuent comme il se doit au total global de la matière. Cela vaut également pour les seuils d'attribution des notes finales fixes pour la composante, qui auront initialement été déterminés à partir du nombre de points maximum pour les composantes.

Figure 59

Exemple d'application d'un coefficient pour une matière comportant trois composantes



Le concept de pondération permet de refléter l'importance relative que l'IB accorde aux éléments évalués, en fonction de la contribution au résultat final qu'il souhaite pour chacun. Ainsi, si une composante sert avant tout à évaluer l'interprétation des données ou des sources et que sa pondération est de 30 %, cela implique que, en comparaison des autres objectifs spécifiques du cours, l'interprétation représente 30 % de ce qui est important. Il est fréquent que plusieurs composantes évaluent les mêmes objectifs spécifiques, ce qui rend ce calcul moins pertinent.

Un aspect secondaire, mais important, de la pondération, est qu'elle permet de déterminer le nombre total de points d'une évaluation qui convient aux tâches et aux critères de notation, plutôt que d'essayer de créer artificiellement un total global.

Il est très important de souligner que les notes d'un candidat n'ont pas à correspondre à la pondération de la composante. L'IB est conscient que les candidats n'ont pas tous les mêmes points forts ; c'est pourquoi il utilise des outils d'évaluation très divers. Il est important de comparer les pondérations réelles des résultats de la cohorte entière à celles indiquées lors de la conception de la matière. Si, au cours d'une session, la contribution d'une composante est beaucoup plus forte que prévu, cela peut indiquer que l'épreuve était particulièrement facile et donc qu'il faut chercher d'autres preuves pour l'attribution des notes finales. En revanche, si la contribution d'une composante est régulièrement beaucoup plus forte que prévu, un réexamen de la conception est nécessaire, car cela indique que l'évaluation ne fonctionne pas comme prévu.

L'approche présentée ne reflète pas les méthodes plus complexes de pondération, de calcul (totalisation) et d'application d'un coefficient décrites, par exemple, par Wood (1991, chapitre 10), mais elle est basée sur les principes solides de l'évaluation critériée et soutient les principes de l'évaluation de l'IB.

Cette approche d'obtention de la note définitive par totalisation implique qu'un candidat peut compenser de mauvais résultats obtenus à une composante par des résultats plus élevés obtenus dans une autre, car seule la note totale a des conséquences sur la note finale. Il s'agit là de la notion incarnée par le « système de compensation », par opposition au « système de maîtrise » qui impose au candidat d'afficher le niveau requis dans toutes les composantes pour se voir attribuer une note finale particulière.

Il convient de souligner que la note finale d'un candidat à une matière est obtenue par la totalisation des notes des différentes composantes, et non des notes finales des composantes. Étant donné que chaque note finale d'une composante représente une gamme de notes, il est fort possible que deux candidats avec les mêmes notes finales par composante se voient attribuer des notes finales différentes pour la matière.

Les contrôles de la qualité de l'attribution des notes finales et les rapports sur la répartition

- L'établissement des seuils d'attribution des notes finales est le résultat d'un ensemble de processus visant à aider l'IB à parvenir à une décision mesurée.
- L'examineur en chef et son équipe d'examineurs superviseurs recommandent des seuils d'attribution des notes finales.
- Il incombe au responsable de matière de l'IB de vérifier que l'attribution des notes finales s'est déroulée dans le respect des principes et des processus de l'IB.
- Les recommandations et les justifications (y compris les données essentielles) sont rassemblées dans un rapport sur la répartition des notes dans la matière rédigé sous 24 heures, qui est examiné minutieusement par l'équipe de direction chargée de l'évaluation de l'IB.

Le processus d'attribution des notes finales a été mis en place afin d'aider les examinateurs à émettre leurs jugements en réduisant au minimum les biais involontaires ou à prendre des décisions concernant des travaux atypiques. Le responsable de matière de l'IB est chargé de superviser ce processus afin de veiller à son respect et d'aider les examinateurs superviseurs à détecter les difficultés qui apparaîtront.

Les recommandations initiales quant aux seuils d'attribution des notes finales sont émises par l'examineur en chef et son équipe d'examineurs superviseurs. Ces recommandations doivent être étayées d'une argumentation solide expliquant pourquoi ces seuils sont les plus appropriés, sur la base de preuves disponibles (et parfois contradictoires).

Le responsable de matière de l'IB portera un regard critique amical sur ce processus, afin de garantir que les recommandations de l'examineur en chef sont mesurées et justifiées.

Ces recommandations et justifications sont alors présentées dans le rapport sur la répartition des notes dans la matière rédigé sous 24 heures. Ce rapport comprend également les données sous-jacentes, comme des changements concernant la cohorte étudiant le cours et les notes finales attribuées avec les seuils proposés en comparaison des années précédentes. Ce rapport rédigé sous 24 heures est ensuite étudié par des membres de la direction de la division de l'évaluation (généralement le responsable en chef du programme et le responsable en chef des principes et pratiques de l'évaluation), qui décideront si les arguments avancés sont suffisamment solides.

Si l'équipe de direction de l'évaluation de l'IB n'est pas satisfaite des recommandations faites, elle expliquera ses inquiétudes à l'examineur en chef et lui demandera de donner la priorité à certains aspects des preuves ou d'approfondir son analyse pour mieux soutenir ses recommandations.

Le chef de la direction de l'évaluation de l'IB décide en dernier ressort des seuils d'attribution des notes finales, à la lumière des recommandations faites par les examinateurs en chef.

L'octroi du diplôme ou du certificat du programme

Le diplôme de l'IB, le certificat du POP et le certificat du PEI ne sont pas décernés à l'issue du processus d'attribution des notes finales. Leur octroi est décidé pendant la préparation de la publication des résultats, selon le processus décrit dans la section « [La préparation de la publication des résultats](#) ».

Les observateurs

L'IB s'est engagé à accroître la transparence de ses processus d'évaluation et à faciliter la compréhension générale de son mode d'attribution des notes finales. Dans le cadre de cet engagement, des enseignants sont invités à observer les réunions de délibérations (les modalités dépendent de la façon dont se tiennent ces réunions, c'est-à-dire de façon traditionnelle ou en ligne). Les observateurs devront rendre compte de leur expérience à leurs collègues et préparer un rapport à l'intention de toute la communauté de l'IBEN (le réseau de collaborateurs de l'IB). Pour obtenir de plus amples informations, veuillez adresser un courriel à support@ibo.org.

Les principes de l'attribution des notes finales

Pour décider de l'attribution des notes finales, l'IB suit un certain nombre de principes fondamentaux.

1. Les seuils d'attribution des notes finales 3 – 4, 6 – 7 et 2 – 3 sont déterminés (dans cet ordre) à partir de toutes les preuves disponibles (appréciations et statistiques). Lorsqu'aucun travail de candidat ne peut servir directement à l'établissement de ces seuils, les preuves disponibles seront utilisées pour établir le seuil le plus approprié.
2. Les autres seuils d'attribution des notes finales sont ensuite déterminés de façon arithmétique, en suivant la procédure adaptée.
3. Les décisions relatives aux seuils d'attribution des notes finales sont prises sur la base d'une analyse triangulaire des preuves issues des jugements des examinateurs, des preuves statistiques et des informations sur la cohorte. Toutes ces preuves doivent peser de façon égale et un compromis doit être trouvé.
4. Si la cohorte passant l'évaluation présente de grandes similitudes avec celles des années précédentes, les résultats devraient être globalement similaires aux précédents. Toutefois :
 - a. les cohortes changent souvent beaucoup d'une année à l'autre, notamment dans les matières présentées par un petit nombre de candidats ;
 - b. si les résultats sont très différents des années précédentes, des preuves solides devront en expliquer les raisons.
5. Si les tâches d'une évaluation sont globalement similaires à celles des années précédentes, les seuils d'attribution des notes finales devraient être similaires aux précédents.
 - a. Tous les seuils d'attribution des notes finales peuvent changer chaque année, y compris pour les tâches d'évaluation interne.
 - b. L'IB s'efforce de s'assurer que les évaluations présentent le même niveau de difficulté chaque année, mais ce niveau pourra varier dans certaines épreuves.
6. La priorité est donnée aux seuils d'attribution des notes finales par matière. C'est ce résultat qui revêt une réelle importance pour le candidat et qui sert de base aux parties prenantes pour prendre les décisions.
7. Les seuils d'attribution des notes finales par composante sont une étape essentielle pour obtenir un résultat global solide par matière, mais l'accumulation d'effets limités au niveau des composantes peut avoir un effet considérable sur les résultats dans la matière prise dans son ensemble.

Les contrôles de la qualité fondés sur les copies dites « à risque »

- Le résultat le plus important de l'évaluation est la note finale que le candidat reçoit. Elle est donc au cœur des vérifications finales.
- Ces vérifications finales visent à mettre en lumière tout élément inhabituel. Toute inquiétude relative à la notation aura été levée avant cette phase.
- Si des modèles de modification des notes sont détectés durant ce processus, l'IB mènera une enquête, et il est possible qu'il soit décidé de recorriger tous les travaux notés par un examinateur.
- Ces vérifications sont réalisées par les examinateurs les plus constants et, en cas de différence, leur note est considérée comme plus appropriée que la précédente.

Le modèle d'assurance de la qualité de la notation passe par des vérifications régulières de la norme de notation des examinateurs de l'IB au moyen des copies de contrôle, afin de s'assurer que les autres copies sont notées correctement. Un autre système permettant de contrôler la qualité consiste à chercher des résultats inattendus et à vérifier leur exactitude. Ce système porte le nom de « vérification des copies à risque ».

Deux critères permettent actuellement de rechercher les résultats inhabituels.

- Un candidat a obtenu une note finale bien plus basse que celle prévue par l'enseignant.
- Les résultats globaux d'un établissement sont très différents de ceux obtenus l'année précédente. L'accent est particulièrement mis sur les cas où une composante montre un schéma très différent des années précédentes.

Les deux cas ci-dessus n'indiquent pas que les notes attribuées sont incorrectes, mais ils mettent en évidence certains domaines dans lesquels un autre contrôle de la qualité peut être nécessaire. En règle générale, la priorité est donnée aux cas pour lesquels il existe des preuves que les candidats ont moins bien réussi que prévu, plutôt que mieux.

Il est uniquement fait appel aux examinateurs qui se sont montrés les plus constants (lors des contrôles prévus par le modèle d'assurance de la qualité), ce qui permet d'être certain de l'exactitude des notes qu'ils attribuent. De ce fait, s'ils proposent une note différente, leur jugement sera utilisé plutôt que celui de l'examineur précédent. Cependant, si une différence considérable est observée entre deux examinateurs constants, un examen plus poussé sera effectué, et il est possible qu'un troisième avis soit demandé pour comprendre cet écart.

Avec la notation électronique, toutes ces notes sont enregistrées par l'ordinateur, mais dans le cas des copies sur papier, les établissements peuvent parfois voir deux ou plusieurs séries de notes du fait de ce processus de révision. L'IB expliquera toujours quelles notes correspondent à l'examineur le plus haut placé dans la hiérarchie et le plus fiable.

La notation des copies à risque poursuit un double objectif :

1. vérifier que les candidats ayant initialement obtenu un résultat inhabituel obtiennent en fin de compte un résultat juste ;
2. chercher des signes d'un problème de notation récurrent (par exemple, une question que plusieurs examinateurs ont eu du mal à noter ou un manque de constance dans la notation de certains examinateurs). Dans de telles situations, les travaux des candidats concernés seront recorrigés.

Le comité d'attribution des notes finales

- Le comité d'attribution des notes finales est l'organe qui prend les décisions stratégiques lors de chaque session d'examens.
- Il est constitué de membres de la direction de plusieurs divisions de l'IB et d'examineurs en chef chargés de différentes disciplines.
- En plus de donner son approbation finale pour la publication des résultats, il étudie également les cas de mauvaise conduite, les demandes de mesures compensatoires et toute autre question pertinente.
- Outre ses membres votants, des établissements scolaires et des interlocuteurs peuvent être invités à observer ses délibérations.

Le comité d'attribution des notes finales est chargé de la dernière étape du processus d'approbation des résultats des candidats. Il s'agit de l'organe qui passe en revue les recommandations du chef de la direction de l'évaluation de l'IB relatives au respect des normes de l'IB pendant la session. Les résultats doivent être attribués par le comité, pour le compte du Conseil de fondation. Ce comité établit également la politique et les précédents qui s'appliquent à l'octroi des titres de l'IB.

La composition et les attributions précises du comité varient légèrement en fonction du programme, mais il est en principe constitué du même nombre de membres votants issus :

- des membres de la direction des divisions de l'évaluation, des études et des services aux établissements de l'IB ;
- des examineurs en chef des différentes disciplines.

Contrairement aux membres de la direction de la division de l'évaluation de l'IB qui effectuent des vérifications dans les rapports sur la répartition des notes rédigés sous 24 heures, le comité d'attribution des notes finales agit comme un organe de surveillance. Il examine les résultats au niveau global (par exemple, les taux de réussite globaux au programme) ainsi que tous les points portés à son attention par les membres du personnel de la division de l'évaluation de l'IB.

De même, le comité d'attribution des notes finales passe en revue les cas de mauvaise conduite et de mauvaise administration. Des recommandations sont faites par un sous-comité, elles sont débattues, et des précédents sont établis dans les nouveaux cas.

Les demandes de mesures compensatoires sont traitées de la même manière : les membres du personnel de la division de l'évaluation de l'IB font des recommandations au comité d'attribution des notes finales, puis celles-ci sont débattues afin d'aboutir à une décision finale.

Le dernier rôle du comité d'attribution des notes finales consiste à mener une réflexion sur la réussite de la session d'examens et à faire des recommandations à l'IB pour les sessions ultérieures.

Les conflits d'intérêts

Le comité d'attribution des notes finales est un organe de prise de décisions important. De ce fait, il est essentiel qu'il donne une image de transparence et d'indépendance. La politique adoptée est donc la suivante : tout membre perçu comme ayant un conflit d'intérêts pour l'un des éléments à l'ordre du jour quittera la pièce pour la discussion concernée, une mesure de précaution qui est reconduite au début de chaque réunion.

Les observateurs

L'IB est très favorable à la présence d'observateurs aux réunions du comité d'attribution des notes finales, et ce, pour les raisons suivantes :

- rendre plus transparentes les procédures suivies par le comité d'attribution des notes finales ;
- permettre un retour d'information sur les procédures suivies par le comité d'attribution des notes finales ;
- reconnaître le partenariat entre les différentes parties constitutives de l'IB ;
- donner la possibilité de faire des suggestions de modification et d'amélioration des procédures suivies par le comité d'attribution des notes finales, au moyen d'un compte rendu écrit.

Pour que l'IB puisse atteindre ces objectifs, il demande aux observateurs de remettre un rapport sur leur participation au chef de la direction de l'évaluation de l'IB dans les deux semaines qui suivent la réunion. Ils doivent y faire des observations d'ordre général sur les procédures et processus actuellement suivis, et suggérer les modifications et améliorations qu'ils jugent nécessaires. Aucun commentaire ne doit être fait sur un cas précis abordé par le comité.

En raison du caractère sensible des questions débattues par le comité, les observateurs ont également des obligations à respecter en matière de confidentialité et de conflit d'intérêts. Les observateurs n'ont pas de droit de vote pour les décisions du comité.

Pour savoir plus précisément comment devenir observateur d'une réunion du comité d'attribution des notes finales, veuillez envoyer un courriel à support@ibo.org.

La préparation de la publication des résultats

- Une fois toutes les copies notées et tous les seuils d'attribution des notes finales établis, plusieurs processus doivent être suivis pour garantir que les résultats des candidats sont prêts à être publiés.
- Lorsqu'il n'existe aucune preuve du travail d'un candidat, l'IB doit estimer la note en question en suivant la procédure pour note manquante.
- Les résultats des candidats doivent être combinés de façon à déterminer s'ils peuvent recevoir le diplôme de l'IB, le certificat du POP ou le certificat du PEI.
- Tous les cas des candidats ayant bénéficié de mesures compensatoires doivent être revus.
- La publication des résultats impose avant tout de mettre en place un système de contrôle des changements efficace, afin que toute modification des notes finales soit détectée et que les interlocuteurs concernés, essentiellement les établissements scolaires et les universités, en soient informés.

L'établissement des seuils d'attribution des notes finales n'est pas directement suivi de la publication des résultats pour les candidats : plusieurs procédures doivent être suivies entre ces deux étapes. Si quelques-unes concernent seulement certains candidats, comme la conservation des notes finales pour les matières anticipées ou la mise en œuvre de mesures compensatoires, d'autres s'appliquent à tous.

Les notes manquantes

- Lorsque l'IB n'est pas en mesure d'accéder au travail d'un candidat pour une raison indépendante de la volonté de l'établissement scolaire ou du candidat, une note sera estimée de façon à minimiser les éventuels effets négatifs sur le candidat.
- Une telle estimation doit se fonder sur des preuves et, si les travaux du candidat ne sont pas suffisants pour être exploitables, une autre solution doit être trouvée pour remplacer la note manquante.
- La procédure pour note manquante se base sur les résultats moyens du candidat. Ainsi, autant de candidats se verront avantagés que désavantagés par cette estimation. Le résultat le plus juste sera toujours celui obtenu en corrigeant le travail réel du candidat.

Il peut arriver que le travail d'un candidat ne puisse être remis à l'IB en vue d'une notation pour des raisons indépendantes de la volonté de l'établissement scolaire et du candidat. C'est notamment le cas lorsque des épreuves d'examen sont perdues par les services postaux ou que des candidats contractent subitement une maladie le jour de l'examen.

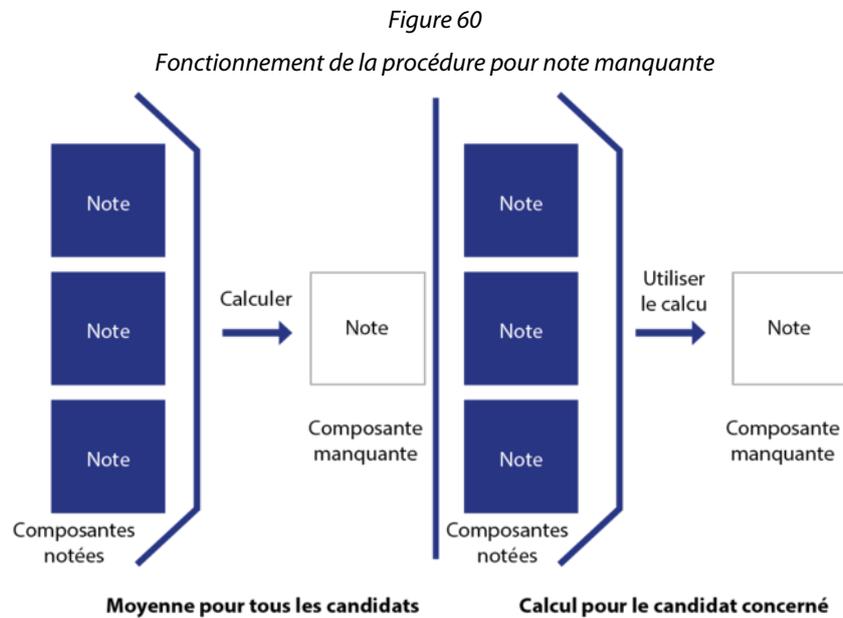
Dans de telles situations, il est possible d'engager la procédure pour note manquante afin d'estimer une note pour le candidat concerné et lui éviter ainsi d'être injustement désavantagé.

Il est pertinent de lancer cette procédure dans les cas où l'absence du travail du candidat est due aux actions de l'IB ou de tiers (à l'exclusion de l'établissement) et où il ne serait pas raisonnable de demander au candidat de recommencer l'évaluation.

La procédure pour note manquante n'a pas vocation à remplacer les aménagements de la procédure d'évaluation à des fins d'inclusion, ni les mesures compensatoires (sauf cas exceptionnel).

Quelle que soit la situation, la procédure pour note manquante doit se fonder sur des preuves des accomplissements du candidat. De ce fait, si les informations sont peu nombreuses, il est très peu probable qu'une note juste puisse être estimée.

La procédure pour note manquante s'inspire des résultats moyens obtenus par tous les candidats dans la composante concernée, qui sont comparés aux résultats obtenus dans les autres composantes. Il est important de reconnaître qu'en utilisant cette moyenne, il est attendu qu'autant de candidats obtiendront de moins bons résultats avec cette estimation que de meilleurs résultats. Par conséquent, le résultat sera toujours plus juste s'il est possible de noter le travail réel du candidat.



S'il n'existe qu'une composante dans le modèle d'évaluation d'une matière, il est impossible d'utiliser la procédure pour note manquante, car aucune preuve ne pourra être exploitée pour estimer la note. Cette situation se présente dans le PEI, ce qui justifie l'adoption d'une approche spécifique à ce programme pour déterminer la note finale manquante.

Les résultats au programme

- L'IB octroie un diplôme ou un certificat dès lors que certains critères précis sont respectés et non sur la base d'un jugement au cas par cas.
- Chaque programme possède ses propres conditions requises pour l'obtention du certificat.

Une fois que les résultats d'un candidat sont connus dans toutes les matières, un calcul est effectué pour déterminer s'il peut prétendre au diplôme de l'IB, au certificat du POP ou au certificat du PEI. Cette décision repose sur la satisfaction d'un ensemble de critères expliqués dans la section consacrée à chaque titre.

Le mode de calcul du résultat au Programme du diplôme

Le nombre total de points obtenus au Programme du diplôme est calculé en ajoutant les notes finales (allant de 1 à 7) attribuées dans chacune des six matières au nombre de points obtenus pour le tronc commun (entre 0 et 3). Cela signifie que le nombre de points maximum est 45*.

Cela signifie que, pour calculer le nombre total de points d'un candidat, une importance égale est accordée à toutes les matières, qu'elles soient étudiées au NM ou au NS.

* Le maximum de 45 points peut être atteint en obtenant 6 (nombre de matières) fois 7 points (note finale maximale) plus 3 points (nombre maximum de points pour le tronc commun).

La matrice des points pour le tronc commun

À la différence des autres matières, l'échelle de notes va de A à E pour la théorie de la connaissance (TdC) et le mémoire. Aucune note finale n'est attribuée pour la troisième composante du tronc commun, à savoir le programme CAS, car il ne serait guère pertinent de noter le travail réalisé par un candidat dans ce domaine. Le tronc commun vaut entre 0 et 3 points dans le nombre total de points pouvant être obtenus au Programme du diplôme. Par ailleurs, un candidat peut échouer au diplôme de l'IB s'il obtient un E en TdC ou au mémoire, ou s'il n'achève pas son programme CAS. Le nombre de points est calculé à l'aide du tableau qui suit.

Figure 65

La matrice des points pour le tronc commun

		Théorie de la connaissance (TdC)					
		Note finale attribuée	A	B	C	D	E
Mémoire	A	3	3	2	2	Facteur éliminatoire	
	B	3	2	2	1		
	C	2	2	1	0		
	D	2	1	0	0		
	E	Facteur éliminatoire					

Les facteurs éliminatoires

Un candidat ne pourra pas se voir décerner le diplôme de l'IB si l'un des neuf cas de figure ci-après est observé.

- Les exigences du programme CAS n'ont pas été satisfaites.
- Le total des points du candidat est inférieur à 24.
- Un « N » (indiquant l'absence d'une note finale) a été attribué pour la théorie de la connaissance, le mémoire ou toute matière concourant au diplôme de l'IB.
- La note finale « E » a été attribuée pour la théorie de la connaissance ou pour le mémoire, ou pour les deux.
- Une note finale de 1 a été attribuée dans une matière ou un niveau.
- Une note finale de 2 a été attribuée dans trois matières ou plus (au NS ou au NM).
- Une note finale égale ou inférieure à 3 a été attribuée dans quatre matières ou plus (au NS ou au NM).
- Le candidat a obtenu moins de 12 points pour les matières au NS (seules les trois notes finales les plus élevées comptent pour les candidats inscrits à quatre matières au NS).
- Le candidat a obtenu moins de 9 points pour les matières au NM (les candidats inscrits à deux matières au NM doivent obtenir au moins 5 points au NM).

Le diplôme bilingue de l'IB

À la place du diplôme de l'IB habituel, un « diplôme bilingue » peut être décerné à un candidat qui :

- présente deux langues du groupe 1 et obtient une note finale égale ou supérieure à 3 pour chacune de ces deux langues ;
- présente l'un des cours du groupe 3 ou 4 dans une langue différente de celle choisie pour le groupe 1. Le candidat doit obtenir une note finale égale ou supérieure à 3 à la fois pour la langue du groupe 1 et pour le cours du groupe 3 ou 4.

Les matières pilotes et les matières interdisciplinaires peuvent contribuer à l'obtention d'un diplôme bilingue, si les conditions ci-dessus sont remplies.

Les éléments suivants ne peuvent contribuer à l'obtention d'un diplôme bilingue :

- le mémoire ;
- un programme propre à l'établissement ;
- une matière choisie par le candidat en complément des six matières du diplôme de l'IB (« matière supplémentaire »).

Le mode de calcul du résultat au POP

Aucun résultat général sous forme de points n'est associé au certificat du POP.

Un candidat se verra décerner le certificat du POP, sous réserve du respect des exigences suivantes :

- l'établissement a confirmé que le candidat a satisfait aux exigences de la formation à orientation professionnelle ;
- le candidat a obtenu une note finale égale ou supérieure à 3 dans au moins deux des cours du Programme du diplôme ;
- le candidat a obtenu une note finale égale ou supérieure à D pour le projet de réflexion ;
- l'établissement a confirmé que le candidat a satisfait à toutes les exigences des autres composantes du tronc commun (cours de compétences personnelles et professionnelles, apprentissage par le service et cours de développement d'une langue seconde) ;
- le candidat n'a pas fait l'objet d'une sanction pour mauvaise conduite de la part du comité d'attribution des notes finales.
- Les notes finales des matières du Programme du diplôme et du projet de réflexion sont confirmées par le même comité d'attribution des notes finales que pour le Programme du diplôme.

Le certificat bilingue du POP

À la place du certificat du POP habituel, un « certificat bilingue » peut être décerné à un candidat qui :

- présente deux cours du groupe de matières Études en langue et littérature du Programme du diplôme dans deux langues différentes et obtient une note finale égale ou supérieure à 3 pour chacun de ces deux cours ;
- présente un cours du groupe de matières Études en langue et littérature du Programme du diplôme ainsi qu'un cours du groupe de matières Individus et sociétés ou Sciences du Programme du diplôme dans une langue d'usage différente de celle choisie pour le groupe Études en langue et littérature et obtient une note finale égale ou supérieure à 3 dans ces deux cours.

Le mode de calcul du résultat au PEI

À la fin de la 5e année du PEI, les élèves peuvent être inscrits aux évaluations externes de l'IB. Les résultats obtenus aux évaluations seront consignés dans un document présentant les résultats de cours du PEI. Par ailleurs, ces candidats peuvent choisir de passer des évaluations qui conduiront à l'octroi du certificat du PEI.

Les établissements peuvent également délivrer une attestation de participation au PEI. Elle s'adresse aux élèves du PEI qui ont suivi le programme pendant au moins 2 ans et satisfait aux exigences en 3^e ou 4^e année du PEI. Ces élèves ne sont en aucune façon inscrits aux évaluations de l'IB. L'attestation de participation est un document propre à l'établissement qui n'est pas vérifié par l'IB.

Pour pouvoir obtenir le certificat du PEI de l'IB, le candidat doit avoir suivi la dernière année du programme, après une participation recommandée de deux ans, et :

- passer un examen sur ordinateur ou réaliser un portfolio électronique dans six matières choisies dans les groupes suivants : Langue et littérature, Acquisition de langues (ou une deuxième matière du

groupe Langue et littérature), Individus et sociétés, Mathématiques, Sciences, et enfin, au choix, Arts, Éducation physique et à la santé ou Design ;

- obtenir une note finale égale ou supérieure à 3 dans chacune des six matières susmentionnées ;
- passer l'examen sur ordinateur pour l'apprentissage interdisciplinaire et obtenir une note finale égale ou supérieure à 3 ;
- achever le projet personnel et obtenir une note finale égale ou supérieure à 3 ;
- obtenir un total général de 28 points ;
- satisfaire aux attentes de l'établissement scolaire concernant le service communautaire.

Pour se voir décerner un certificat bilingue du PEI, le candidat doit en outre obtenir des résultats satisfaisants aux examens sur ordinateur dans l'un des domaines suivants :

- un second cours de Langue et littérature (au lieu d'un cours d'Acquisition de langues) ;
- un cours (ou plus) du groupe de matières Sciences ou Individus et sociétés, ou l'examen interdisciplinaire, dans une langue autre que celle choisie pour le cours de Langue et littérature.

L'équité pour tous : répondre aux besoins des candidats

Pour que les évaluations de l'IB soient valides, aucune discrimination ne doit être pratiquée à l'encontre des candidats ayant des besoins particuliers. À cet effet, l'IB étudiera toute demande de modification des épreuves d'examen et d'aménagement de la procédure d'évaluation à des fins d'inclusion, comme il est expliqué dans le document *Procédures d'évaluation* spécifique à chaque programme.

- Le meilleur moyen de garantir l'équité est de concevoir une évaluation accessible à tous, qui supprime de ce fait tout besoin de modification. Cette idée est englobée dans la notion de conception universelle de l'évaluation, qui fait partie intégrante de l'engagement de l'IB envers la conception universelle de l'apprentissage (CUA).
- Lorsque les besoins des candidats sont connus en avance, ils peuvent donner lieu à des aménagements de la procédure d'évaluation à des fins d'accès et d'inclusion, parmi lesquels la modification des épreuves d'examen.
- D'autres circonstances se présentent à courte échéance ou ne peuvent être gérées au moyen des aménagements de la procédure d'évaluation à des fins d'inclusion. Le cas échéant, des processus de mesures compensatoires sont alors engagés.
- Enfin, le but de tous ces aménagements est de créer une équité pour tous les candidats. Par conséquent, lorsque l'IB doit prendre une décision, il doit tenir compte de ce qui est juste pour l'ensemble de la cohorte et non uniquement pour le candidat concerné. L'objectif est de parvenir à une égalité des chances pour tous les candidats.

L'IB estime que tous les candidats doivent pouvoir passer leurs examens dans des conditions aussi équitables que possible. Il reconnaît que les conditions normales d'évaluation peuvent désavantager les candidats ayant des besoins en matière de soutien à l'apprentissage en les empêchant de montrer le niveau qu'ils ont atteint. De même, il admet que, parfois, des événements ou des circonstances qui échappent au contrôle des candidats auront une incidence sur leurs résultats et donc qu'ils doivent être pris en considération.

Le meilleur moyen de garantir l'équité d'une évaluation est de faire en sorte que tout le monde passe la même évaluation de la même manière. Nombre des modifications apportées pour répondre à des besoins spécifiques pourraient aider tous les candidats à comprendre les questions et à les traiter. Dans l'idéal, l'IB devrait concevoir toutes ses évaluations sur la base d'une compréhension des différents besoins des candidats au lieu de traiter certains candidats différemment. Cette idée est celle de la conception universelle de l'évaluation. L'IB a conscience que cette approche prônant une inclusion totale ne peut pas toujours être adoptée ; c'est pourquoi il est également possible de demander des aménagements de la procédure d'évaluation à des fins d'inclusion.

Les **aménagements à des fins d'accès et d'inclusion** sont prévus pour les candidats qui présentent des besoins individuels tels que :

- des troubles d'apprentissage ;
- des difficultés du langage ;
- des difficultés spécifiques d'apprentissage ;
- des difficultés de communication et des troubles de la parole ;
- des troubles du spectre autistique ;
- des troubles sociaux, émotionnels et comportementaux ;
- des handicaps multiples, des troubles physiques, des troubles sensoriels, des problèmes médicaux ou de santé mentale.

Tous les aménagements raisonnables seront considérés pour un candidat présentant des besoins spécifiques. Pour obtenir de plus amples informations, veuillez consulter le document *Procédures d'évaluation* spécifique à votre programme et la publication de l'IB intitulée *Candidats ayant des besoins en matière d'aménagement de la procédure d'évaluation*.

Les **circonstances défavorables** sont celles échappant au contrôle du candidat et risquant d'avoir des effets négatifs sur ses résultats. Ces circonstances sont étudiées par le comité d'attribution des notes finales et, en cas d'issue favorable, les candidats près du seuil d'attribution des notes finales se verront attribuer la note finale supérieure.

Le principe d'équité pour tous les candidats reconnu par l'IB implique que l'étude des circonstances défavorables et des demandes d'aménagements de la procédure d'évaluation à des fins d'inclusion ne crée pas de situation injuste pour les autres candidats passant les évaluations. L'objectif est de parvenir à une égalité des chances pour tous les candidats.

Principes sous-tendant les aménagements à des fins d'accès et d'inclusion

Les principes qui sous-tendent les aménagements de la procédure à des fins d'accès et d'inclusion sont exposés dans le document de l'IB intitulé *Candidats ayant des besoins en matière d'aménagement de la procédure d'évaluation*. Le texte ci-dessous est extrait de la version consacrée au Programme du diplôme, mais des principes similaires s'appliquent aux autres programmes.

1. L'IB doit garantir qu'une note finale attribuée dans une matière reflète bien le niveau d'accomplissement d'un candidat. Par conséquent, les mêmes normes d'évaluation s'appliquent à tous les candidats, qu'ils aient ou non des besoins en matière de soutien à l'apprentissage.
2. Les aménagements à des fins d'accès et d'inclusion, y compris les ajustements raisonnables, sont des mesures prises en amont des examens pour que l'évaluation soit à la portée du candidat concerné. Ils ne peuvent pas être demandés rétrospectivement, que ce soit pour les épreuves orales ou écrites.
3. Les aménagements demandés pour un candidat ne doivent cependant pas lui offrir un avantage dans une composante d'évaluation.
4. Les aménagements à des fins d'accès et d'inclusion décrits dans le présent document sont destinés aux candidats qui possèdent l'aptitude à satisfaire à toutes les modalités d'évaluation permettant l'octroi du diplôme ou des résultats de cours du Programme du diplôme.
5. Si des aménagements de la procédure à des fins d'accès et d'inclusion sont nécessaires à un candidat pendant ses études dans le cadre du Programme du diplôme ou lors de ses examens d'entraînement, l'établissement est à même de les mettre en place. En revanche, si des aménagements sont requis aux fins de l'évaluation, le présent document dresse la liste de ce qui peut se faire sans l'autorisation préalable de l'IB. Dans toutes les autres situations, il est impératif d'obtenir l'autorisation préalable du centre mondial de l'IB à Cardiff. De la même manière, lorsqu'un candidat du Programme du diplôme a des difficultés pour répondre aux exigences du programme créativité, activité, service (CAS), le service L'IB vous répond doit être consulté.
6. Il est conseillé aux établissements de prévoir des aménagements à des fins d'accès et d'inclusion pour leurs candidats sur la base des critères énoncés dans la présente politique et des observations des enseignants quant au comportement des candidats pendant les travaux et les examens réalisés en classe.
7. Les aménagements à des fins d'accès et d'inclusion demandés pour un candidat doivent correspondre à sa méthode de travail habituelle dans le cadre de ses études. En ce qui concerne les aménagements ne correspondant pas à la méthode de travail habituelle du candidat et n'ayant été mis en place pour soutenir le candidat qu'au cours des six derniers mois de sa scolarité ou ultérieurement, à savoir juste avant les examens, l'IB autorisera de telles demandes d'aménagements à des fins d'accès et d'inclusion uniquement dans des cas très exceptionnels et particuliers.
8. L'IB souhaite accorder des aménagements à des fins d'accès et d'inclusion qui soient compatibles avec ceux normalement en vigueur pour le candidat concerné. Toutefois, seuls les aménagements qui sont

en accord avec la politique et les pratiques de l'IB seront autorisés. Il ne faut pas partir du principe que l'IB acceptera nécessairement les aménagements demandés par un établissement. Les coordonnateurs doivent fournir des informations sur les méthodes de travail habituelles du candidat en classe.

9. L'IB est attaché à une philosophie pédagogique fondée sur la sensibilité internationale. Par conséquent, sa politique en matière d'aménagements à des fins d'accès et d'inclusion ne reflète pas l'usage établi d'un pays particulier. Elle résulte de l'analyse des pratiques en vigueur dans différents pays, pour offrir les mêmes chances aux candidats ayant des besoins en matière d'aménagement de la procédure d'évaluation.
10. L'IB s'efforcera d'offrir, dans la mesure du possible, les mêmes aménagements à des candidats présentant le même type de besoins. Compte tenu des différences culturelles durant la reconnaissance des besoins en matière de soutien à l'apprentissage et de la nature des aménagements de la procédure d'évaluation autorisés dans les établissements, il est parfois nécessaire de trouver des compromis pour permettre une comparaison entre les candidats de différents pays.
11. Toute demande d'aménagements à des fins d'accès et d'inclusion sera jugée au cas par cas. Une autorisation antérieure d'aménagements, par l'IB ou par un autre organisme décernant des titres, n'influencera pas la décision relative à l'autorisation des aménagements demandés par le coordonnateur.
12. L'IB traite toutes les informations relatives à un candidat avec confidentialité. Seuls le personnel de l'IB habilité et les membres du comité d'attribution des notes finales y ont accès si nécessaire et sont tenus de respecter cette confidentialité.
13. Lorsqu'un établissement ne respecte pas les conditions définies par l'IB dans l'octroi des aménagements à des fins d'accès et d'inclusion ou lorsqu'il en accorde sans l'aval de l'IB, le candidat risque de ne pas recevoir de note finale de l'IB pour la matière et le niveau concernés.
14. S'il est possible de prouver que le faible niveau d'un candidat dans sa langue d'usage est dû à un besoin en matière de soutien à l'apprentissage identifié, des aménagements à des fins d'accès et d'inclusion peuvent être accordés. (Pour les matières des groupes 3 à 6, tous les candidats ont le droit d'utiliser un dictionnaire bilingue lors des épreuves écrites.)
15. Les établissements ne doivent pas informer les examinateurs des circonstances défavorables ou des problèmes affectant un candidat (tels que l'autisme, des difficultés rédactionnelles, etc.).
16. Pour le travail évalué en interne, les enseignants ne doivent apporter aucun ajustement dans leur notation.
17. La liste des aménagements à des fins d'accès et d'inclusion disponibles est régulièrement mise à jour. L'IB examinera d'autres aménagements suggérés par un coordonnateur, à condition qu'ils puissent être mis à disposition de tous les autres candidats ayant les mêmes besoins.

Le document intitulé *Règlement général du Programme du diplôme* stipule qu'un candidat peut se présenter à trois sessions d'examens pour obtenir le diplôme. Un candidat ayant des besoins en matière de soutien à l'apprentissage peut bénéficier de sessions supplémentaires, à la discrétion de l'IB.

1. Lorsque la nature du trouble d'un candidat ou la nature des aménagements à des fins d'accès et d'inclusion autorisés risquent de gêner les autres candidats pendant un examen, le candidat en question doit passer l'épreuve dans une pièce séparée et être surveillé conformément aux réglementations régissant le déroulement des examens du Programme du diplôme.
2. Les épreuves écrites doivent être surveillées conformément aux réglementations régissant le déroulement des examens du Programme du diplôme. La personne surveillant l'examen du candidat ne doit pas être un membre de sa famille ni quelqu'un avec qui il pourrait y avoir un conflit d'intérêts apparent ou perçu comme tel.
3. Le service L'IB vous répond doit être informé dans les meilleurs délais de toute difficulté liée à la nature des aménagements à des fins d'accès et d'inclusion ou de toute difficulté imprévue éprouvée par le candidat pendant les examens.

Les dispenses d'évaluation

Les dispenses ne sont normalement admises pour aucune composante d'évaluation. Cependant, lorsqu'une composante d'évaluation ou une partie d'une composante d'évaluation demande une fonction physiologique qu'un candidat ne possède pas, ce dernier peut en être dispensé. Toutes les possibilités d'ajustement raisonnable doivent avoir été étudiées avant de déposer une demande de dispense pour une composante d'évaluation. Une autorisation de dispense ne sera accordée que pour des motifs suffisants. L'incapacité physique d'un candidat à réaliser les fonctions demandées pour la composante doit être clairement documentée, et de façon exhaustive.

Pour connaître le détail des principes et des processus liés aux dispenses d'évaluation, veuillez vous reporter au document intitulé *Candidats ayant des besoins en matière d'aménagement de la procédure d'évaluation* correspondant au programme qui vous intéresse.

Les possibilités d'aménagements de la procédure d'évaluation à des fins d'inclusion pour l'évaluation sur ordinateur

Le recours à l'évaluation sur ordinateur permet au candidat d'avoir un plus grand contrôle sur la manière dont l'évaluation est présentée. Les ordinateurs permettent en effet de proposer tout un éventail de polices de caractères, de tailles de texte et de couleurs pour répondre aux besoins précis de chacun, et ce type d'ajustement peut facilement être mis à la portée de tous les candidats.

Dans la plupart des demandes récentes, les aménagements de la procédure d'évaluation à des fins d'inclusion visaient l'autorisation d'utiliser un ordinateur, un besoin qui est à l'évidence comblé par l'évaluation électronique, à condition que le logiciel permettant l'inclusion soit compatible avec l'outil à l'écran. L'IB a parfaitement conscience de cette exigence et fait le nécessaire pour s'assurer que toutes les évaluations sur ordinateur respecteront les normes industrielles associées à de tels logiciels permettant l'inclusion.

Les circonstances défavorables

Les circonstances défavorables ou imprévues sont celles échappant au contrôle du candidat et susceptibles d'être préjudiciables à ses résultats. Ce sont, par exemple, une maladie ou blessure temporaire, un stress grave, des circonstances familiales particulièrement éprouvantes, un deuil ou des événements pouvant menacer la santé ou la sécurité du candidat. Les circonstances défavorables peuvent également inclure des événements qui affectent la communauté scolaire dans son ensemble, comme des troubles civils ou une catastrophe naturelle.

Les circonstances défavorables n'incluent pas les insuffisances dues au fait de l'établissement scolaire, notamment les problèmes liés au personnel enseignant. Il incombe à l'établissement de s'assurer que tous les candidats respectent les exigences du programme et de l'évaluation.

Pour connaître le détail de ce qui est inclus dans la catégorie des circonstances défavorables et ce qui en est exclu, veuillez vous reporter au document *Procédures d'évaluation* et au *Règlement général* du programme qui vous intéresse.

Pour déterminer si un candidat affecté par des circonstances défavorables peut bénéficier de mesures compensatoires, le comité d'attribution des notes finales étudiera les preuves fournies par l'établissement. Si les circonstances qui affectent le candidat sont considérées comme « défavorables », et que le candidat est autorisé à bénéficier de mesures compensatoires, les notes totales du candidat pourront être ajustées dans les matières ou les composantes du programme concernées. Si un ou deux points séparent la note du candidat du seuil d'attribution des notes finales supérieur, sa note finale pour les matières concernées sera augmentée.

La conception universelle de l'évaluation

Les paragraphes précédents s'intéressaient aux situations pouvant amener des candidats à demander des modifications ou une aide particulière pour passer les évaluations de l'IB équitablement, ainsi qu'à la manière dont l'IB gérait ces situations. La meilleure solution reste toutefois d'avoir des évaluations ne dressant pas d'obstacle à la participation. La conception universelle de l'évaluation consiste à prendre en considération l'accès, l'inclusion, l'égalité, les sensibilités culturelles, les stéréotypes et les biais dès le début de la conception de l'évaluation. Elle inclut la création des tâches d'évaluation et des questions d'examen, mais elle remonte également à la conception du modèle d'évaluation global, qui détermine la manière de créer des évaluations comparables chaque session. Si des évaluations plus inclusives sont créées dès le départ, non seulement la pertinence de construit est accrue, mais encore les risques de ne pas satisfaire les besoins de tous les candidats sont minimisés.

La conception universelle de l'évaluation est un aspect de la notion plus globale de conception universelle de l'apprentissage (CUA). La CUA vise à créer des environnements d'apprentissage accessibles à tous les apprenants, notamment aux candidats ayant une déficience, aux candidats de langues ou de cultures différentes, et aux candidats doués et talentueux. Les principes de la CUA peuvent être appliqués à tous les domaines de l'éducation, notamment la conception du programme d'études, la gestion de l'établissement scolaire et l'enseignement. Pour en savoir plus sur la CUA dans le contexte de l'IB, vous pouvez consulter l'ouvrage suivant (en anglais) : RAO, K., CURRIE-RUBIN, R., et LOGLI, C. 2016. *UDL and Inclusive Practices in IB Schools Worldwide*.

La publication des résultats

Du point de vue de l'IB, la publication des résultats consiste essentiellement à mettre en œuvre un protocole très strict de contrôle des modifications.

Au cours de la session d'examens, les informations sont constamment mises à jour, ce qui permet d'avoir en permanence une idée précise de ce qui se passe. Une fois les résultats portés à la connaissance des candidats et des établissements scolaires, il doit leur être indiqué très clairement qu'aucune modification ne se produira sans que toutes les parties prenantes en soient dûment informées. En ce qui concerne le Programme du diplôme et le POP, cela vaut tout particulièrement pour les universités, qui recevront un relevé de notes rassemblant tous les résultats du candidat.

Des modifications peuvent être autorisées après la publication des résultats en raison de réclamations concernant les résultats, d'une confirmation des résultats en suspens par les établissements scolaires ou de la résolution de problèmes d'intégrité intellectuelle.

Les réclamations concernant les résultats, la procédure d'appel et le retour d'information général

- Les réclamations concernant les résultats visent à permettre aux établissements d'indiquer à l'IB où ils estiment qu'une erreur a été commise au cours du processus de notation.
- Les normes de notation appliquées pendant les recorections effectuées dans le cadre des réclamations concernant les résultats doivent être strictement les mêmes que celles suivies lors de la session d'examens principale.
- Pour s'en assurer, l'IB s'appuie sur les contrôles prévus par le modèle d'assurance de la qualité. Les évaluations externes de l'IB se veulent davantage sommatives que formatives ; c'est pourquoi il est demandé aux examinateurs de ne rédiger des commentaires que s'ils expliquent leur notation.
- L'IB a également une procédure d'appel officielle que les établissements ou les candidats peuvent engager s'ils estiment que les processus n'ont pas été correctement suivis.

Pour consulter une description des aspects juridiques et des procédures des réclamations concernant les résultats proposées par l'IB, veuillez vous référer au document *Procédures d'évaluation* spécifique au programme qui vous intéresse.

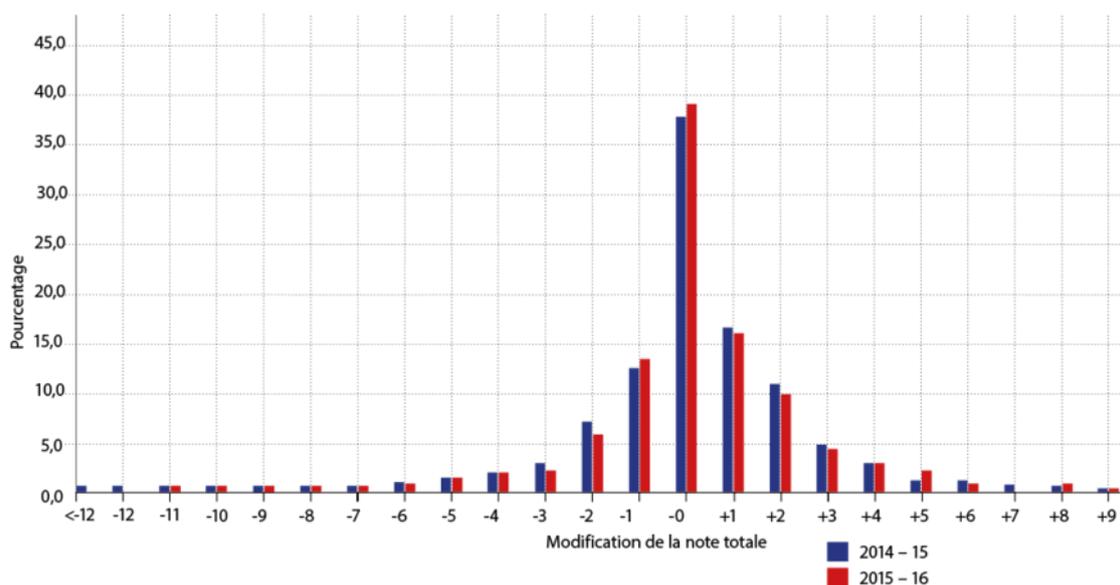
Le but des réclamations concernant les résultats

Le processus de réclamation concernant les résultats constitue le dernier rempart contre les erreurs dans le système de notation. Il permet aux établissements scolaires d'indiquer à l'IB s'ils estiment qu'une erreur a été commise, et à l'IB d'effectuer les vérifications nécessaires et d'apporter les corrections éventuellement requises.

Comme cela a été indiqué dans la section sur le modèle d'assurance de la qualité de la notation, il est possible que deux examinateurs ne soient pas tout à fait d'accord sur le nombre de points à attribuer à une question, sans qu'aucun d'eux n'ait commis d'erreur. Par conséquent, si le processus de réclamation concernant les résultats aboutit à une modification mineure du nombre de points, il est probable qu'aucune erreur n'ait été commise lors de la notation. Et c'est le cas dans la plupart des demandes de réclamation concernant les résultats.

Figure 61

Changements apportés aux notes après des réclamations concernant les résultats en mai 2015 et mai 2016



Dans le processus de réclamation concernant les résultats, l'IB fait appel uniquement aux examinateurs les plus constants et les plus haut placés dans la hiérarchie.

Étant donné que le système repose sur des seuils d'attribution des notes finales, il est toujours possible que des candidats se trouvent un point au-dessous d'un seuil ou au contraire juste au-dessus. Dans de tels cas, une réclamation concernant les résultats pourrait aboutir à un changement de note finale, malgré l'absence d'erreur dans la notation indiquée précédemment, et l'IB soutiendrait que le candidat se trouve à la limite de deux notes finales et que toutes deux sont un reflet fidèle de son accomplissement.

Pour réduire les effets de ces différences mineures, mais appropriées, de notation entre les examinateurs, tous les examens externes du candidat sont recorrectés lors d'une réclamation concernant les résultats. Cela limite en effet l'impact d'un changement mineur, car, en supposant que les changements ne sont pas systématiques, ils devraient s'annuler sur plusieurs épreuves.

Le maintien des normes

Les réclamations concernant les résultats ont pour objectif de donner aux candidats la note qu'ils auraient dû obtenir durant la session de notation. Il est toutefois indispensable que les examinateurs ne se laissent pas influencer par l'insatisfaction d'un candidat à l'égard de sa note initiale ou par le fait que cette note est proche d'un seuil d'attribution des notes finales.

Pour aider les examinateurs à respecter la norme, l'IB glisse des copies de contrôle dans les travaux visés par une réclamation concernant les résultats. Les correcteurs reçoivent ainsi une alerte électronique dès qu'ils s'éloignent de la norme convenue.

Pour empêcher tout examinateur de commettre une grave erreur au cours d'une recorection résultant d'une réclamation concernant les résultats, l'IB vérifie tous les changements de notes proposés à l'issue de cette procédure. Ainsi, dès qu'une différence considérable et inexplicée est observée entre deux examinateurs, un troisième avis sur la copie est demandé. Dans ce type de situation, la note attribuée par l'examineur le plus haut placé dans la hiérarchie sera généralement retenue.

La mise en lumière des problèmes récurrents

L'IB contrôle tous les changements de notes décidés au cours de la procédure de réclamation concernant les résultats, afin de vérifier si certains schémas indiquent un problème avec les notes originales, comme un examinateur qui a réussi les contrôles de la qualité mais qui n'a pas constamment respecté la même norme de notation.

S'il existe des preuves de ces problèmes récurrents, elles sont étudiées pour déterminer si une cause profonde peut être établie. En cas de problème de cet ordre, tous les candidats concernés (qu'ils aient déposé une réclamation concernant les résultats ou non) seront recherchés, afin de vérifier qu'ils n'ont pas été désavantagés.

Il est également fréquent que des problèmes récurrents soient signalés par des établissements. Leurs inquiétudes sont prises très au sérieux, et toutes les preuves disponibles sont étudiées, mais, sauf exception, une recorection ne sera pas engagée si elle n'a pas été demandée par l'établissement scolaire au moyen de la procédure de réclamation concernant les résultats.

Les catégories de réclamations concernant les résultats

Après la publication des résultats, le coordonnateur peut soumettre une demande pour les éléments suivants.

- Réclamation de catégorie 1 : recorection de toutes les composantes d'un candidat évaluées en externe pour une matière donnée
- Rapport de recorection de catégorie 1 : rapport sur la notation réalisée dans le cadre d'une réclamation concernant les résultats de catégorie 1
- Réclamation de catégorie 2 : retour des copies des composantes évaluées en externe
- Réclamation de catégorie 3 : nouvelle révision de notation d'une composante évaluée en interne

Des frais sont facturés pour chacune des catégories ci-dessus (à l'exception des cas où une note finale est modifiée à la suite d'une recorection de catégorie 1).

Une copie renvoyée peut contenir des commentaires utiles de l'examineur à qui la copie a été attribuée. Cependant, l'ajout de ces commentaires ne peut être garanti puisque les examinateurs ne sont pas tenus de rédiger des commentaires dans le cadre de leur travail de notation des travaux de candidats. Cela s'explique par le fait que l'évaluation sommative a pour objectif de présenter une image fidèle de l'accomplissement d'un candidat à la fin de ses études et que l'important pour les examinateurs est de garantir une bonne notation et non de fournir des remarques ou recommandations sur le travail lui-même. Les commentaires sur les réalisations du candidat constituent un élément important de l'évaluation formative menée par l'enseignant tout au long du cours.

Comme il est indiqué dans la section [Révision de notation](#) en ce qui concerne les travaux évalués en externe, la note attribuée par l'examineur et figurant sur la copie ne correspond pas nécessairement à la note définitive « révisée ». La note attribuée au candidat sur IBIS est la bonne.

L'équité pour tous (augmentation et diminution des notes finales)

- À l'issue d'une réclamation de catégorie 1, les notes peuvent être augmentées ou au contraire diminuées. Il est donc impératif que les candidats donnent leur autorisation pour que leur travail passe par cette procédure.
- L'IB ne peut pas raisonnablement exiger d'un établissement qu'il obtienne l'autorisation de tous les candidats d'une classe avant de demander une réclamation de catégorie 3 (nouvelle révision de notation). Par conséquent, dans ce cas (uniquement), les notes finales sont protégées et ne peuvent pas être diminuées.

- Si la note d'un candidat venait à être modifiée pour une autre raison, la nécessité de protéger la note finale serait décidée au cas par cas. Cependant, si aucune faute n'a été commise par le candidat ou par l'établissement, il ne sera en principe pas autorisé de diminuer la note finale.

Le système de réclamation concernant les résultats a pour objectif de veiller à ce que les candidats reçoivent la bonne note. L'IB doit donc faire le nécessaire pour garantir que la note issue de cette procédure est aussi exacte que possible. Pour ce faire, il est demandé aux examinateurs les plus haut placés dans la hiérarchie et les plus constants de se charger de la notation en cas de réclamation concernant les résultats et de revoir leurs résultats pour s'assurer qu'ils peuvent justifier leur décision.

Avec de telles mesures de précaution, il est considéré que la note décidée à l'issue d'une réclamation concernant les résultats est le reflet le plus fidèle du travail d'un candidat, d'où l'attribution de cette note et de cette note finale. Cela implique que le résultat d'un candidat peut diminuer ou au contraire augmenter du fait d'une réclamation de catégorie 1. Les établissements doivent donc impérativement obtenir l'autorisation du candidat avant d'envoyer une réclamation de ce type.

Les établissements doivent donc impérativement obtenir l'autorisation du candidat avant d'envoyer une réclamation de ce type. Pour ce qui est des réclamations de catégorie 3 (nouvelle révision de notation), tout changement apporté aux notes pourrait s'appliquer à la totalité de la cohorte. Si le principe de l'obtention de l'autorisation des candidats était observé, cela signifierait qu'en ne donnant pas son autorisation ou en ne répondant pas à la demande, un seul candidat pourrait empêcher tous les autres candidats de sa classe de voir leur travail être de nouveau révisé. Par conséquent, pour que les notes finales des candidats ne soient pas diminuées sans qu'ils aient officiellement accepté ce risque, les notes finales peuvent uniquement être augmentées (et jamais diminuées) à l'issue d'une réclamation de catégorie 3.

Si la note d'un candidat venait à être modifiée pour une autre raison, la nécessité de protéger la note finale serait décidée au cas par cas. Cependant, si aucune faute n'a été commise par le candidat ou par l'établissement, tout laisse supposer que la note finale sera protégée, car le candidat aura pu prendre certaines décisions irrévocables sur la base de ce résultat.

Le retour d'information issu des évaluations notées

Les évaluations qui sont évoquées dans cette section sont les évaluations sommatives. Bien que rien ne s'oppose à leur utilisation à des fins formatives, l'IB doit prendre soin de s'assurer que leur validité n'est pas altérée par d'autres emplois.

Il est attendu des examinateurs qu'ils procèdent à une notation correcte et appropriée, dans le respect de la norme établie par l'examineur principal. Il leur est donc uniquement demandé d'inclure des commentaires pour expliquer leur notation, et non de faire un retour d'information aux candidats pour leur indiquer comment s'améliorer. Comme le savent les enseignants, fournir un retour d'information instructif fait appel à des compétences différentes de celles utilisées lors de la notation sommative, et cela ne peut se faire convenablement, en principe, sans avoir une connaissance globale du candidat. Cela est en totale contradiction avec l'idée de dégager les examinateurs de tout biais induit par les candidats dans leur notation.

Le simple fait de pouvoir réfléchir sur ce qui a été bien accompli dans une évaluation peut aider les candidats et les enseignants à améliorer leurs compétences et leur compréhension ; c'est pourquoi l'IB apporte volontiers des renseignements sur les résultats de la notation, en précisant cependant que ces commentaires n'ont pas vocation à être formatifs.

La procédure d'appel

La procédure de réclamation concernant les résultats doit apporter la garantie que toutes les erreurs relatives à la notation ou aux processus d'évaluation ont été rectifiées. Si un établissement ou un candidat n'est toujours pas convaincu que ses évaluations ont été notées correctement, il peut se pourvoir en appel. La procédure d'appel n'est toutefois envisageable qu'une fois toutes les autres voies de recours épuisées.

Il est ainsi possible de faire appel pour les motifs suivants :

1. résultats : lorsqu'un établissement scolaire a des raisons de croire que les résultats d'un candidat ne sont pas corrects à l'issue de toutes les procédures appropriées en matière de réclamation concernant les résultats ;
2. décision confirmant la mauvaise conduite : action ne contestant pas la sévérité de la sanction ;
3. décision concernant une mesure compensatoire : action faisant suite à un refus d'accorder une mesure compensatoire à un candidat en raison de circonstances défavorables présumées ;
4. décision administrative non couverte par les circonstances susmentionnées affectant les résultats d'un ou de plusieurs candidats.

La procédure d'appel comporte deux niveaux. Au premier niveau, un membre de la direction de l'IB n'étant pas encore intervenu dans le cas examinera les preuves et la décision. Si le plaignant n'est pas satisfait de l'issue du premier niveau, il peut demander le deuxième niveau d'appel. Les preuves et la décision seront alors examinées par un comité constitué de trois membres, dont une personne indépendante de l'IB.

Des informations détaillées concernant la procédure d'appel sont disponibles dans le Règlement général spécifique à chaque programme. Pour toute autre information, vous pouvez envoyer un courriel à appeals@ibo.org ou vous mettre en rapport avec le service [L'IB vous répond](#).

L'IB dispose également d'un médiateur qui « travaille avec des individus et des groupes au sein d'une organisation afin de les aider à trouver des solutions en cas de conflit ou de problème ». Les établissements ou les candidats qui ne sont pas satisfaits du processus d'évaluation ont la possibilité de se tourner vers lui.

La préparation des évaluations de l'année suivante

- L'évaluation est considérée comme un cycle. Il est donc important de tirer des enseignements de chaque session pour améliorer constamment les évaluations.
- Les épreuves et les barèmes de notation sont élaborés jusqu'à deux ans avant une session. Divers éléments seront donc pris en compte à différents moments du cycle de vie de l'évaluation.
- Les preuves à prendre en compte vont des expériences opérationnelles aux résultats des candidats, en passant par le retour d'information des enseignants.
- L'IB tient à ce que toute la gamme d'outils disponibles soit utilisée pour comprendre et rectifier tout problème de comparabilité entre les matières, et pas simplement pour apporter des ajustements aux seuils d'attribution des notes finales.
- Toute modification susceptible d'avoir des conséquences sur la méthode d'enseignement devra être apportée sur une durée équivalente à la durée du cours (c'est-à-dire deux ans).

L'utilisation des preuves

L'autoréflexion est un aspect essentiel de la philosophie de l'IB, et elle est tout aussi importante dans le processus d'évaluation que pour les candidats qui sont évalués. Chaque session est l'occasion de recueillir des preuves qui montrent :

- la manière dont les candidats ont interprété les questions d'examen et la correspondance avec les attentes initialement nourries ;
- si chaque question a rempli les attentes de son auteur (Les exigences et le niveau de difficulté de la question ont-ils été correctement évalués ?) ;
- le niveau de fiabilité de la notation des examinateurs pour chaque question (A-t-il été particulièrement difficile de respecter la norme pour certaines questions ?) ;
- si les différentes composantes ont fonctionné comme prévu (Les pondérations réelles correspondaient-elles à celles souhaitées ?) ;
- les problèmes rencontrés lors de la préparation ou de la notation des examens ;
- si les enseignants ont eu le sentiment que l'épreuve répondait à leurs attentes (Cela n'a-t-il pas été le cas pour une question en particulier ?) ;
- si les enseignants ont eu le sentiment que l'épreuve était plus facile ou plus difficile que les épreuves des sessions précédentes (Pour quelles raisons ?).

Ces informations seront utilisées de différentes façons, en fonction de la nature du problème. Si le problème concerne les questions en cours de préparation, le style sera généralement adapté pendant la phase de rédaction des examens (jusqu'à deux ans avant qu'ils soient passés par les candidats), à moins que sa gravité exige une réécriture des examens préparés.

À l'inverse, lorsque les preuves laissent apparaître que les processus d'évaluation ou les normes de notation ont besoin d'un ajustement, ce dernier pourra être mis en place pour la session suivante.

La comparabilité entre les matières

La comparabilité constitue l'un des aspects les plus difficiles de la validité. Comme le montre l'illustration ci-après, il est souvent difficile de déterminer à quel moment il est raisonnable de vouloir comparer deux choses différentes, et la comparaison des matières est tout aussi complexe. Par exemple, comment comparer des résultats en littérature et en mathématiques ?

Figure 62

Quels éléments est-il raisonnable de comparer ? L'approche adoptée est-elle relative ou absolue ?



Cette situation se voit encore compliquée par le fait que les candidats sont meilleurs dans certaines matières que dans d'autres. De ce fait, il est impossible de se servir d'un candidat « moyen » pour déterminer la difficulté d'une évaluation. De très nombreux chercheurs ont tenté de mesurer la comparabilité entre les matières. C'est notamment le cas dans l'étude de l'Ofqual (2015) intitulée *Inter-Subject Comparability: A Review of the Technical Literature: ISC Working Paper 2*.

Dans les programmes de l'IB, les élèves doivent généralement étudier un programme d'études vaste, en choisissant leurs matières dans différents groupes. Il est donc fondamental de veiller à ce que les matières composant chaque groupe soient aussi exigeantes les unes que les autres, afin de garantir que toutes les voies possibles présentent un niveau de difficulté équivalent. Cela veut par exemple dire veiller à rendre la biologie et la physique, ou la géographie et l'histoire, comparables.

Les comparaisons entre les groupes de matières sont elles aussi importantes, afin de réserver le même traitement aux candidats, quels que soient leurs points forts.

L'IB veille avant tout à accorder autant de poids à chacune des formes de preuves dans ses prises de décisions.

Dans certaines matières, notamment les langues, les examinateurs peuvent comparer le niveau de difficulté en comparant les travaux réalisés dans chacune d'entre elles. Cette comparaison a généralement lieu dans le cadre du processus de normalisation.

Après chaque session, l'IB procède à une analyse des paires de matières. Cette méthode bien documentée (Nuttall *et al.*, 1974) consiste à s'intéresser aux candidats qui ont choisi les deux matières en question et à comparer les notes finales obtenues dans chacune. Pour chaque matière, il devient alors possible de calculer une différence moyenne avec toutes les autres matières et d'établir un classement de la difficulté. Cette méthode a ses limites (voir Ofqual, 2015 ou Coe *et al.*, 2008), mais combinée à d'autres preuves, elle constitue une source précieuse d'informations.

Lorsque l'analyse des paires de matières ou les avis des spécialistes des matières semblent indiquer une différence de niveau entre les matières, cela doit être débattu lors de l'attribution des notes finales. Par exemple, si une matière est qualifiée de « difficile » à l'issue de l'analyse des paires de matières, il faut s'attendre à ce que les examinateurs se montrent généreux dans leurs appréciations. L'IB tient à ce que toute la gamme d'outils disponibles soit utilisée pour comprendre et rectifier tout problème de comparabilité entre les matières, et pas simplement pour apporter des ajustements aux seuils d'attribution des notes finales.

L'introduction des modifications

Avant d'apporter des modifications à une évaluation, il est important de s'assurer que cela ne désavantagera ni les candidats ni les enseignants. Autrement dit, si un remaniement risque d'avoir des conséquences sur les pratiques d'enseignement ou sur la préparation aux examens, il n'interviendra que pour les candidats qui commencent le cours.

Une aide pour la révision du programme

Les preuves tirées du bilan de l'évaluation sont utilisées non seulement pour améliorer les futures évaluations, mais aussi pour faciliter le processus de révision du programme. Un rapport officiel est rédigé sur l'évaluation dans le cadre de chaque révision du programme, mais les responsables des programmes d'études de l'IB participeront souvent à la première réunion d'attribution des notes finales d'un nouveau cours, afin d'observer directement les expériences vécues par les examinateurs.

Les évaluations sur ordinateur

L'évaluation sur ordinateur ne doit en rien être différente de l'évaluation traditionnelle pour ce qui est des enseignements tirés des évaluations des sessions précédentes. L'ordinateur ouvre la voie à des questions innovantes et originales, ce qui constitue un atout majeur. Un examen minutieux de ces questions devra toutefois être réalisé, afin de s'assurer qu'elles satisfont aux attentes.

Les commentaires destinés aux établissements

- L'évaluation sommative de l'IB a pour objectif de mesurer les accomplissements des candidats, et tous les processus doivent être pensés de manière à rendre les résultats les plus valides possible.
- Cela passe notamment par la transparence, afin que les candidats et les enseignants puissent constater que les notes et les notes finales ont été attribuées avec la même constance et sans biais.
- Cet aspect est particulièrement important pour l'évaluation interne, car la note finale est attribuée après une révision de la notation de l'enseignant.
- L'évaluation sommative de l'IB, et notamment le processus de réclamation concernant les résultats, ne vise pas à indiquer aux établissements comment améliorer les résultats de leurs candidats, à moins que cela ne soit une conséquence indirecte des données requises pour une évaluation valide.
- En dehors du processus d'évaluation, les établissements peuvent bénéficier d'autres services de l'IB en matière de méthode d'enseignement et de perfectionnement professionnel.

Les commentaires des examinateurs

On ne peut pas s'occuper de deux choses à la fois. C'est comme les citrouilles dans l'eau : l'une refait surface tandis que l'on essaie de maintenir l'autre sous l'eau.

(Proverbe chinois)

L'évaluation sommative de l'IB a pour objectif de mesurer les accomplissements des candidats. Il est donc demandé aux examinateurs de se concentrer exclusivement sur la notation du travail des candidats, en respectant la norme requise. Ils ne doivent inclure des commentaires que si cela les aide dans leur travail de notation.

Ils ne doivent inclure des commentaires que si cela les aide dans leur travail de notation. Pour rédiger des commentaires constructifs à l'intention des candidats ou des enseignants, l'examineur doit déterminer la note qui correspond au travail, puis tenter d'expliquer de quelle manière ce travail aurait pu être amélioré. La seconde tâche n'est pas insignifiante, car elle constitue l'une des caractéristiques essentielles d'un bon enseignement. Elle requiert cependant du temps et une réflexion qui éloigneront l'examineur de sa mission principale, à savoir corriger le travail selon une norme de notation cohérente. En termes simples, l'IB attend des examinateurs qu'ils effectuent une seule tâche (notation) en maintenant un niveau élevé de qualité, plutôt que deux tâches (notation et retour d'information) avec un niveau de qualité moindre.

L'examineur ne peut émettre un jugement qu'à partir du travail mis à sa disposition, alors que les enseignants expérimentés pourront se baser sur tout un éventail d'informations pour décider du retour d'information à faire au candidat. Le retour de l'examineur sera donc moins complet que celui de l'enseignant.

Pour toutes ces raisons, l'IB indique très clairement à ses examinateurs qu'ils doivent uniquement noter le travail du candidat en respectant la norme, et non ajouter des commentaires pour faire un retour d'information au candidat ou à l'enseignant.

Les examinateurs doivent indiquer clairement à quels endroits les points ont été attribués et, en cas d'ambiguïté, expliquer ce choix en ajoutant des commentaires appropriés. Cela permet à l'IB de vérifier les normes de notation et rend le processus d'attribution des points plus transparent pour les établissements.

Le rapport de recorection de catégorie 1, qui peut être demandé par les établissements à l'occasion d'une réclamation concernant les résultats, constitue la seule exception à ce principe de rédaction des commentaires pour expliquer la notation uniquement. Dans ce cas précis, un examinateur haut placé dans la hiérarchie répondra aux préoccupations exprimées par l'établissement dans le cadre de la réclamation

concernant les résultats, et ce, avec un niveau de détail plus poussé que ce qui est généralement attendu des examinateurs.

Les rapports pédagogiques

Chaque examinateur en chef est tenu de rédiger un rapport pédagogique à l'issue de chaque session. Ce rapport vise à fournir des informations aux enseignants sur l'accomplissement de tous les candidats ayant participé à la session, et notamment des informations sur les questions et les thèmes qui ont été particulièrement bien ou mal traités. Ce rapport donne des renseignements sur chaque composante, traite de la qualité globale des réponses des candidats et inclut des recommandations d'ordre général sur les améliorations qui peuvent être apportées pour les sessions ultérieures. Beaucoup de rapports pédagogiques fournissent également des données clés, tirées des commentaires des enseignants qui ont été pris en compte lors de l'attribution des notes finales. Ces rapports comprennent enfin les seuils d'attribution des notes finales de la matière et les seuils d'attribution des notes finales de la composante qui ont été utilisés pour parvenir aux résultats globaux.

Les enseignants devront évidemment replacer ce retour d'information dans le contexte spécifique de leur classe. Par exemple, même si une question donnée a globalement produit des réponses médiocres, il est tout fait possible que tous les candidats de leur classe aient obtenu des notes élevées.

Les commentaires sur l'évaluation interne

La révision de notation de l'évaluation interne vise à garantir que tous les enseignants suivent la même norme de notation, et l'IB aimerait pouvoir éviter d'avoir à appliquer des facteurs de révision de notation aux travaux d'un établissement. Pour que les enseignants puissent comprendre où ils se situent par rapport à la norme globale, l'IB fournit un retour d'information sur la notation des tâches d'évaluation interne. Cela permet aux enseignants de comprendre les raisons de l'application d'un facteur de révision de notation.

Le retour d'information qui est fait aux établissements ne vise pas à expliquer comment les candidats de l'échantillonnage auraient pu obtenir de meilleurs résultats.

Remarques sur les manuels, les ateliers et les examens

Le programme d'études de chaque cours de l'IB est détaillé dans le guide pédagogique correspondant et sert de base pour la conception des évaluations. Tous les manuels, y compris ceux approuvés par l'IB, visent à aider les enseignants et les apprenants à étudier le cours conformément à la description qui en est faite dans le guide pédagogique, et non à définir le contenu du programme d'études.

Par conséquent, les questions d'examen pourront porter sur un thème ou une partie d'un thème qui figure dans le guide pédagogique mais pas dans un manuel donné. Tous les enseignants sont donc encouragés à se référer uniquement aux guides pédagogiques pour tout ce qui concerne le contenu du programme d'études. Dans le cadre du processus de rédaction des examens, les manuels fréquemment utilisés sont passés en revue, afin de vérifier que les questions d'examen envisagées ne figurent pas déjà dans ces ressources.

De même, les commentaires et les documents fournis dans les ateliers animés par l'IB ne se substituent pas aux énoncés figurant dans les guides pédagogiques, même s'ils peuvent aider les enseignants à mieux interpréter les formulations trouvées dans les guides. Toutes les modifications apportées aux guides pédagogiques seront officiellement publiées par l'IB et clairement indiquées comme telles.

En quoi consistent les processus spécifiques aux programmes ?

- La grande idée derrière les évaluations sommatives externes des programmes de l'IB est de doter les élèves d'un « passeport » qui les aidera à poursuivre leurs études ou à intégrer le monde du travail.
- Étant donné que les quatre programmes de l'IB reposent sur la même philosophie pédagogique, toutes les évaluations de l'organisation doivent suivre les mêmes principes et adopter les mêmes pratiques générales pour respecter cette philosophie. Les objectifs de l'évaluation évolueront naturellement avec la progression des élèves à travers les différents programmes et en fonction des stades de maturité scolaire correspondants.
- Même si l'évaluation est un élément important du PP, ce programme ne se prête pas à une évaluation sommative externe.
- Chaque programme a des particularités qui imposent des processus spécifiques s'inscrivant dans les pratiques de l'évaluation.

Les raisons de l'inclusion d'une évaluation sommative externe dans les programmes de l'IB

Il a été décidé d'inclure une évaluation sommative externe dans le Programme du diplôme, le POP et le PEI (sous forme d'option en dernière année) afin de doter les élèves ayant suivi les programmes de l'IB d'un « passeport » qui les aidera à poursuivre leurs études ou à intégrer le monde du travail.

L'IB se doit d'organiser cette évaluation sommative externe afin de garantir que le « passeport » proposé aux élèves de l'IB respecte la mission et la philosophie de l'organisation, dont les principaux éléments sont résumés ci-dessous.

- Priorité donnée à une éducation globale reposant sur un programme d'études, et non à des matières restreintes
- Valorisation de la sensibilité internationale, à l'aide du profil de l'apprenant
- Priorité donnée à des situations d'évaluation riches de sens et authentiques, et non à une évaluation procédurale reposant sur le rappel de faits
- Autorisation d'une différenciation équitable entre les élèves dans des contextes appropriés
- Recherche d'un effet de remous positif sur l'enseignement et l'apprentissage

Toutes ces valeurs sont résumées dans les principes de l'évaluation de l'IB.

Les élèves qui suivent le PP ont bien moins besoin d'un tel « passeport », d'où l'absence d'évaluation sommative évaluée en externe par l'IB. Les principes d'une bonne évaluation qui sous-tendent la philosophie de l'IB sont toutefois respectés.

L'objectif secondaire de l'évaluation au sein de l'IB est de permettre à l'organisation de montrer la voie en donnant l'exemple dans un monde qui accorde une grande valeur aux évaluations.

Enfin, l'évaluation sommative permet de vérifier que l'établissement comprend de la même manière que l'IB ce qu'est une éducation de qualité, et de le confirmer à l'établissement lui-même, mais aussi aux enseignants, aux parents et aux élèves.

Des besoins et des solutions spécifiques aux programmes

Dans les sections précédentes, les processus qui constituent le cycle d'évaluation ont été expliqués, et ces pratiques généralisées valent pour tous les programmes de l'IB. Chaque programme poursuit cependant un objectif particulier, ce qui explique que des différences puissent être observées dans certains aspects de l'évaluation (par exemple, la façon dont le mémoire et le projet personnel sont présentés et évalués, ainsi que les règles permettant de déterminer si un diplôme ou un certificat peut être octroyé pour un programme).

Ces approches de l'évaluation, qui sont spécifiques à chaque programme, sont importantes. Cette section permettra d'expliquer en quoi elles consistent et comment elles sont gérées.

Passer de la conception du programme d'études à l'évaluation pour revenir à la conception du programme d'études

Dans les deux sections précédentes, il a été expliqué que l'évaluation devait être intégrée dans les objectifs globaux du cours. Par voie de conséquence, le modèle d'évaluation doit faire partie intégrante du développement du cours et l'influencer pour que les résultats de l'évaluation reflètent de manière efficace et juste ce que le cours est censé produire.

L'IB prévoit dans son approche de la révision des programmes d'études des matières une discussion sur l'évaluation, et certaines ressources pour les enseignants mises à la disposition des établissements incluent des exemples montrant la forme que prendront les prochaines évaluations ainsi que des exemples de travaux de candidats annotés pour expliquer la notation ou l'évaluation.

Les exemples de matériel d'examen, comme les spécimens d'épreuves, ont pour objectif :

- de montrer la structure des examens, c'est-à-dire quand les candidats doivent répondre à toutes les questions d'une section, quand un choix leur est proposé ou la longueur de chaque section et épreuve et les points correspondant à chacune ;
- de donner une indication du style des questions et des types de documents servant de stimulus dans l'examen, notamment pour les nouveaux éléments d'un cours ;
- de fournir le contenu indicatif d'une épreuve d'examen réelle, notamment pour comprendre quels éléments du résumé du programme (ou du programme d'études) sont imposés ou facultatifs ;

de fournir aux enseignants un matériel qui leur permettra de préparer leurs candidats aux examens **de manière raisonnable** (par exemple, en organisant un examen d'entraînement).

Les spécimens d'épreuves peuvent contenir des exemples de questions portant sur tous les aspects du cours, ou même sur les nouveaux aspects d'un cours. Ces spécimens doivent être conçus de manière à pouvoir être utilisés efficacement comme de « vrais » examens de l'IB, et non être élaborés à des fins purement illustratives.

Le socle commun à tous les programmes

- Chacun des programmes de l'IB propose un programme d'études ou un cadre pédagogique qui se veut vaste, équilibré, conceptuel et connexe.
- Toutes les évaluations de l'IB doivent être conçues en tenant compte de ces caractéristiques du système éducatif de l'IB, quand bien même elles ne les évaluent pas explicitement, afin qu'il y ait un effet de renouveau positif sur l'enseignement et l'apprentissage.
- Les éléments essentiels qui relient tous les programmes de l'IB sont :
- le profil de l'apprenant ;
- les approches de l'enseignement et de l'apprentissage ;
- la sensibilité internationale.

Figure 63

La continuité entre les programmes de l'IB



Dans le document intitulé *En quoi consiste le système éducatif de l'IB ?*, l'accent est mis sur le fait que les programmes de l'IB favorisent l'apprentissage conceptuel, en se concentrant sur des idées structurantes et marquantes, qui sont pertinentes dans toutes les disciplines et qui aident à assimiler ce qui a été appris et à rendre le programme d'études plus cohérent. Les programmes de l'IB soulignent l'importance d'établir des liens, d'explorer les relations entre les matières scolaires, et d'apprendre des choses sur le monde, et ce, en allant au-delà du cadre de chaque matière. Ils permettent aux élèves d'accéder à un éventail de sujets d'étude et d'expériences d'apprentissage qui se veulent vastes, équilibrés, conceptuels et connexes.

- Vaste et équilibré – Le système éducatif de l'IB offre une approche équilibrée qui permet aux élèves d'accéder à une large gamme de contenus dans de nombreuses matières.
- Conceptuel – L'apprentissage conceptuel se concentre sur de grandes idées structurantes et marquantes, qui sont pertinentes dans toutes les disciplines. Ces idées transcendent les frontières nationales et les barrières culturelles. Les concepts aident à assimiler l'apprentissage, à donner une plus grande cohérence au programme d'études, à approfondir la compréhension propre à la discipline et à renforcer la capacité à examiner des idées complexes. De plus, ils permettent d'utiliser ce qui a été appris dans de nouveaux contextes.

- Connexe – Les cadres pédagogiques de l'IB accordent de l'importance à la simultanéité des apprentissages. Les élèves étudient de nombreuses matières simultanément durant leur parcours scolaire ; ils apprennent à établir des liens et acquièrent une compréhension solide des corrélations entre la connaissance et l'expérience dans nombre de domaines d'étude. Les objectifs globaux des cours et les exigences spécifiques aux programmes permettent véritablement d'acquérir des connaissances sur le monde en allant au-delà des limites traditionnelles de chaque matière.

Dans cette optique, tous les programmes comportent un projet de fin d'études, qui est évalué. Il s'agit de l'exposition pour le PP, du projet personnel ou du projet communautaire pour le PEI, du mémoire pour le Programme du diplôme et du projet de réflexion pour le POP.

Lorsqu'ils conçoivent des évaluations, les enseignants, les établissements et les auteurs de l'IB doivent réfléchir à ces objectifs sous-jacents afin d'éviter de créer des tâches faisant obstacle à un bon enseignement et apprentissage. L'objectif doit systématiquement être de produire un effet de remous positif.

Vecteurs de la mission de l'IB, les programmes de qualité englobent trois domaines :

1. le profil de l'apprenant ;
2. les approches de l'enseignement et les approches de l'apprentissage ;
3. la sensibilité internationale et la compréhension interculturelle.

Pour que les évaluations de l'IB atteignent les objectifs plus généraux des programmes, et soient de ce fait adaptées à l'objectif visé (valides), elles doivent également tenir compte des éléments ci-dessus, même s'ils ne sont pas l'objet même visé par l'évaluation.

Tous les aspects du profil de l'apprenant ne peuvent pas être évalués au moyen d'une évaluation sommative, mais plusieurs sont englobés dans la notion de compétences de pensée de haut niveau. Une bonne évaluation reconnaît l'importance de ces caractéristiques et, même si elle ne vise pas à mesurer ces compétences, elle peut donner aux élèves l'occasion de les développer. Il pourrait s'agir, par exemple, d'encourager des approches éthiques (intègres) des enquêtes et des expériences, en facilitant une révision par les pairs appropriée et en présentant des contextes inattendus aux élèves.

Pour de plus amples informations sur l'approche globale adoptée par l'IB en matière de compétences des élèves, veuillez vous reporter aux documents disponibles sur son [site Web](#) ou à la publication *Des principes à la pratique* spécifique à chaque programme.

Les approches de l'enseignement et les approches de l'apprentissage

Avec ses approches de l'apprentissage et ses approches de l'enseignement, l'IB ne souhaite pas imposer de prescriptions ou de restrictions, mais donner des conseils et faire des suggestions. Il ne fait aucun doute que les enseignants et les établissements doivent disposer de suffisamment de liberté pour être créatifs, mais il est indispensable d'orienter les discussions et de mettre en avant les bonnes pratiques. Pour l'IB, l'enseignement des compétences sera efficace si les approches de l'apprentissage sont à la fois explicitement définies et mises en œuvre durablement.

L'amélioration des compétences passe par une consolidation sur le long terme et dans des contextes variés. Quelle que soit la méthode que les enseignants ou les établissements décident de suivre pour intégrer les approches de l'apprentissage dans leurs classes, l'IB considère que ce choix doit être laissé aux personnes qui connaissent le mieux les besoins des élèves, à savoir les enseignants.

Pour plus d'informations à ce sujet, veuillez vous référer aux publications de l'IB consacrées aux approches de l'enseignement et de l'apprentissage ou aux sections pertinentes dans la publication *Des principes à la pratique* spécifique au programme qui vous intéresse.

La sensibilité internationale et la compréhension interculturelle

Les programmes de l'IB sont étudiés par des élèves situés dans de nombreux pays et représentant toute une palette de nationalités. L'IB entend les amener non seulement à atteindre les objectifs pédagogiques de ses programmes, mais aussi à développer « la sensibilité nécessaire pour contribuer à bâtir un monde meilleur et plus paisible, dans un esprit d'entente mutuelle et de respect interculturel » et à « comprendre que les autres, en étant différents, puissent aussi être dans le vrai » (extraits de la déclaration de mission de l'IB, 2002). L'enseignement prôné par l'IB se caractérise donc à la fois par un contexte international et par un objectif de compréhension interculturelle, deux éléments qui doivent se retrouver dans l'évaluation.

Pour y parvenir, le plus important est de faire appel à des spécialistes de l'éducation, notamment des auteurs d'épreuves d'examen et des personnes chargées du développement des programmes d'études, issus de contextes culturels très divers. La mission de l'IB tient à ne pas occulter les différences, mais au contraire à les aborder de façon à ce que les élèves les explorent sans se trouver défavorisés.

Dans certaines disciplines, il est possible de favoriser la diversité culturelle, en tenant compte des différentes perspectives culturelles dans le programme d'études. La biologie, la chimie, la psychologie et les arts visuels offrent des exemples de cette approche. Dans les trois premiers cours, les options permettent aux établissements scolaires de choisir un contenu qui respectera, dans une certaine mesure, leurs traditions culturelles concernant l'enseignement de la matière.

Dans d'autres disciplines, la sensibilité internationale est encouragée au moyen du matériel et des sources d'inspiration que l'élève est incité à utiliser. C'est notamment le cas des matières artistiques et des cours de langue et littérature, mais cette notion peut également intervenir dans un éventail plus large de tâches d'évaluation interne.

La sensibilité internationale ne se réduit pas au fait de connaître et de comprendre d'autres cultures. Les attitudes affichées et les actions entreprises en sont d'autres aspects fondamentaux. Il est difficile d'évaluer

les attitudes à l'aide des évaluations scolaires traditionnelles, qui se concentrent davantage sur les accomplissements que sur les qualités affectives.

L'IB parvient à le faire par l'intermédiaire des éléments non évalués de ses programmes, comme le programme créativité, activité, service dans le Programme du diplôme et le projet communautaire dans le PEI. Les candidats ne peuvent se voir attribuer le diplôme de l'IB ou le certificat du PEI s'ils ne terminent pas avec succès ces composantes, qui, même si elles ne sont pas directement évaluées, influent grandement sur le résultat global du processus d'évaluation de l'IB.

Figure 30

Action fondée sur des principes



Le fait de permettre aux candidats de choisir les questions auxquelles ils souhaitent répondre peut être considéré comme la meilleure façon de répondre aux différentes exigences internationales en matière d'évaluation, mais cela pose certains problèmes du point de vue de la comparabilité entre les options. Cela est toujours le cas lorsque plusieurs questions sont proposées ou que les tâches d'évaluation sont très ouvertes. Même la définition d'une « difficulté égale » est difficile à donner lorsque les candidats sont issus

de milieux scolaires très différents. En règle générale, il est plus facile de garantir la comparabilité avec des tâches communes qui permettent aux candidats de s'appuyer sur leurs expériences personnelles pour rédiger leurs réponses. Il revient alors à l'examineur de réussir à maintenir une norme commune, mais ce défi est moins grand que celui de noter selon la même norme deux tâches qui peuvent présenter des niveaux de difficulté différents.

Il est possible d'obtenir des renseignements sur la comparabilité en analysant les résultats des candidats. Cette analyse est examinée plus en détail dans la section « [L'attribution des notes finales \(et totalisation\)](#) ».

L'évaluation effectuée dans un contexte international présente des défis supplémentaires (en matière d'équité), qui sont plus importants que ceux normalement rencontrés dans un système national. Des questions qui sont parfaitement appropriées dans un pays peuvent devenir complètement inappropriées dans un autre. Les questions portant sur les sports, les voyages, les divertissements, les événements historiques, et même sur la météo, doivent être préparées avec beaucoup de soin. Il semblerait que la seule façon de contourner ce problème soit de rédiger des questions d'examen dénuées de tout contexte socioculturel, à l'exception du plus petit dénominateur commun. Toutefois, cela rendrait non seulement les questions d'examen très limitées et ennuyeuses mais irait également à l'encontre de la philosophie de l'évaluation de l'IB et des bonnes pratiques de l'évaluation, en ce sens que cela ne permettrait pas de garantir la validité au moyen de tâches basées sur le contexte. Pour que l'apprentissage soit bon, il est vital que le travail et l'évaluation tiennent compte du contexte.

Il existe deux façons de résoudre ce dilemme. Premièrement, des informations sur le contexte peuvent être fournies aux candidats, grâce aux précisions apportées dans le contenu du programme de la matière, en leur donnant des études de cas sur lesquelles les questions sont basées, ou encore dans la question d'examen elle-même (à condition que ces informations ne soient pas trop longues et qu'elles ne distraient donc pas les candidats de l'objet de l'évaluation).

La seconde méthode consiste à utiliser des tâches d'évaluation et des questions d'examen plus ouvertes, qui permettent aux candidats de choisir leur propre contexte. Dans cette méthode, la notation doit se concentrer sur une compréhension plus profonde plutôt que sur une simple connaissance du contenu du programme de la matière, car il n'y aura aucune base commune en ce qui concerne le contenu. Cela s'accorde parfaitement avec les principes de l'IB en matière d'évaluation des candidats.

Figure 31

Gamme de normes culturelles et contextes



Quand bien même ces deux méthodes seraient utilisées, les candidats pourraient avoir à réaliser des tâches d'évaluation dont le contexte n'est pas familier dans leur milieu socioculturel. Une fois de plus, cela s'accorde avec la philosophie de l'IB en matière d'évaluation, en ce sens que l'un des objectifs des programmes est de rendre les élèves plus ouverts aux autres façons de faire les choses, plus conscients du monde et plus aptes à fonctionner dans un milieu culturel qui ne leur est pas familier. L'une des exigences concernant la pensée de haut niveau est d'être capable d'utiliser ses connaissances dans des situations inconnues. Il est donc parfaitement approprié d'inclure de tels éléments dans l'évaluation pourvu qu'ils aient un effet égal sur les élèves issus de différents milieux culturels.

Un nombre important d'élèves de l'IB passent les examens dans une langue qui n'est pas leur meilleure langue. Presque tous ces cas concernent l'anglais, car les élèves qui travaillent en français ou en espagnol (les deux autres langues principales utilisées pour les évaluations de l'IB) sont en général des locuteurs natifs. Il faut donc porter une grande attention à la façon dont sont formulées les questions de manière à ne

pas désavantager les élèves qui ne sont pas des locuteurs natifs. Ces précautions sont prises au moment de la révision des épreuves.

L'évaluation sommative de l'IB, tout comme la grande majorité des systèmes d'évaluation officiels, est très individualiste. Comme l'a souligné Brown (2002), cela est en grande partie dû au fait que le Programme du diplôme appartient à une tradition propre à l'Europe occidentale et que les sociétés qui la composent sont par nature individualistes. Les candidats sont presque uniquement évalués sur ce qu'ils accomplissent seuls. On peut considérer cela comme étant culturellement inéquitable, car, dans un certain nombre de cultures, la contribution de l'individu est toujours moins importante que celle du groupe ; dans ces cultures, c'est ce qu'accomplit le groupe qui est important. Il est également vrai qu'en matière d'équité individuelle, il y a des individus qui travaillent mieux en équipe que seuls, et vice-versa. De plus, il est courant, à la fois dans la salle de classe et dans le monde du travail, que les individus travaillent de manière interdépendante plutôt que de manière indépendante.

Lectures supplémentaires

Pour obtenir de plus amples informations sur les valeurs qui sous-tendent le système éducatif de l'IB, veuillez consulter les ressources suivantes :

- *En quoi consiste le système éducatif de l'IB ?*
- *Site Web consacré aux approches de l'enseignement et de l'apprentissage*
- [Profil de l'apprenant de l'IB](#)
- Les guides indiquant comment passer des principes à la pratique pour chacun des programmes (PEI, Programme du diplôme et POP).

Le Programme du diplôme de l'IB

Les caractéristiques de l'évaluation sommative externe du Programme du diplôme sont les suivantes :

- les candidats doivent étudier un nombre prescrit de matières pour obtenir le diplôme de l'IB ;
- le résultat global au diplôme de l'IB est décrit à l'aide d'un total de points sur 45 (maximum) ;
- les composantes du tronc commun peuvent contribuer à hauteur de trois points au résultat global au diplôme de l'IB, selon une matrice des points ;
- presque toutes les matières présentent plusieurs composantes d'évaluation, et combinent évaluation externe et évaluation interne ;
- presque toutes les matières sont proposées au niveau moyen (NM) et au niveau supérieur (NS), et contribuent à parts égales au résultat global au diplôme ; le nombre de candidats inscrits varie considérablement d'une matière à l'autre.

Les objectifs globaux du Programme du diplôme

La validité des résultats de l'évaluation ne peut être déterminée qu'en ayant une idée précise de la finalité des cours et du programme. Cette section commence donc par une présentation des objectifs globaux du programme.

Le Programme du diplôme propose aux élèves âgés de 16 à 19 ans une expérience éducative vaste, équilibrée, stimulante et axée sur la réalité internationale. Les élèves doivent étudier simultanément six matières et les composantes du tronc commun du programme, et ce, pendant deux ans. Le programme est conçu pour doter les élèves des compétences scolaires fondamentales qui leur seront nécessaires pour leurs études universitaires, la poursuite de leur éducation et la profession de leur choix. Il soutient en outre le développement des valeurs et des aptitudes requises pour mener une vie épanouie et pleine de sens.

(Le Programme du diplôme : des principes à la pratique, 2015, p. 7)

Ce qui fait la spécificité du Programme du diplôme (et, en réalité, de tous les programmes de l'IB), c'est l'intérêt porté à l'expérience scolaire intégrale de chaque élève, et cela transparaît dans le choix de l'IB de se concentrer sur la validité du programme dans son ensemble et pas seulement de chaque cours.

Les utilisations valides des résultats des évaluations du Programme du diplôme

Lorsque l'IB élabore ses modèles d'évaluation et ses programmes d'études, il s'efforce de faire en sorte que les notes finales obtenues dans les cours du Programme du diplôme et le total des points obtenus au diplôme puissent être utilisés pour :

- la sélection en vue de l'entrée à l'université ou dans le monde du travail ;
- déterminer si les candidats ont déjà satisfait aux exigences d'un programme universitaire (attribution de crédits ou dispense de certains modules, cours, etc.).

Lorsqu'un candidat a passé une évaluation dans une langue (d'usage), cela fournit également des preuves de sa capacité à poursuivre ses études dans la matière dans cette langue.

Pour comprendre les raisons pour lesquelles il est important de tenir compte de ces utilisations valides, veuillez vous reporter à la section « [Qu'est-ce qui est valide : l'évaluation ou son utilisation ?](#) ».

La structure du Programme du diplôme

Pour se voir octroyer le diplôme de l'IB, tout candidat doit étudier six matières et satisfaire aux exigences des composantes du tronc commun : la théorie de la connaissance, le mémoire et le programme créativité, activité, service (CAS).

Les matières doivent être choisies dans les groupes de matières suivants :

- Études en langue et littérature ;
- Acquisition de langues ;
- Individus et sociétés ;
- Sciences ;
- Mathématiques ;
- Arts.

Au lieu d'un cours choisi dans le groupe de matières Arts, les candidats peuvent étudier une deuxième matière du groupe Études en langue et littérature, Acquisition de langues, Individus et sociétés, Sciences ou Mathématiques.

Certaines matières sont étudiées au niveau moyen (NM) et d'autres au niveau supérieur (NS). Les cours du NS et du NM diffèrent de par leur étendue, mais ils sont évalués en fonction des mêmes descripteurs des notes finales. En revanche, les candidats du NS doivent montrer un ensemble de connaissances ainsi qu'une compréhension et des compétences plus variés.

Trois matières au moins et quatre au plus doivent être présentées au NS, tandis que les matières restantes sont présentées au NM.

Il existe également un nombre restreint de cours interdisciplinaires relevant de plusieurs groupes de matières. Par exemple, le cours de systèmes de l'environnement et sociétés permet de satisfaire aux exigences des groupes de matières Individus et sociétés et Sciences. Avec ces cours interdisciplinaires, les candidats bénéficient d'une plus grande flexibilité pour choisir leurs six matières.

Figure 64

Le modèle du Programme du diplôme : une illustration de la structure du programme



Pour garantir que les élèves ne choisiront pas des matières dont les contenus se recoupent, des règles supplémentaires sont imposées. Elles sont détaillées dans le document *Procédures d'évaluation* publié pour le Programme du diplôme.

Le mode de calcul du résultat au Programme du diplôme

Le nombre total de points obtenus au Programme du diplôme est calculé en ajoutant les notes finales (allant de 1 à 7) attribuées dans chacune des six matières au nombre de points obtenus pour le tronc commun (entre 0 et 3). Cela signifie que le nombre de points maximum est 45*.

Cela signifie que, pour calculer le nombre total de points d'un candidat, une importance égale est accordée à toutes les matières, qu'elles soient étudiées au NM ou au NS.

* Le maximum de 45 points peut être atteint en obtenant 6 (nombre de matières) fois 7 points (note finale maximale) plus 3 points (nombre maximum de points pour le tronc commun).

La matrice des points pour le tronc commun

À la différence des autres matières, l'échelle de notes va de A à E pour la théorie de la connaissance (TdC) et le mémoire. Aucune note finale n'est attribuée pour la troisième composante du tronc commun, à savoir le programme CAS, car il ne serait guère pertinent de noter le travail réalisé par un candidat dans ce domaine.

Le tronc commun vaut entre 0 et 3 points dans le nombre total de points pouvant être obtenus au Programme du diplôme. Par ailleurs, un candidat peut échouer au diplôme de l'IB s'il obtient un E en TdC

ou au mémoire, ou s'il n'achève pas son programme CAS. Le nombre de points est calculé à l'aide du tableau qui suit.

Figure 65

La matrice des points pour le tronc commun

		Théorie de la connaissance (TdC)				
Mémoire	Note finale attribuée	A	B	C	D	E
	A	3	3	2	2	Facteur éliminatoire
	B	3	2	2	1	
	C	2	2	1	0	
	D	2	1	0	0	
	E	Facteur éliminatoire				

Les facteurs éliminatoires

Un candidat ne pourra pas se voir décerner le diplôme de l'IB si l'un des neuf cas de figure ci-après est observé.

- Les exigences du programme CAS n'ont pas été satisfaites.
- Le total des points du candidat est inférieur à 24.
- Un « N » (indiquant l'absence d'une note finale) a été attribué pour la théorie de la connaissance, le mémoire ou toute matière concourant au diplôme de l'IB.
- La note finale « E » a été attribuée pour la théorie de la connaissance ou pour le mémoire, ou pour les deux.
- Une note finale de 1 a été attribuée dans une matière ou un niveau.
- Une note finale de 2 a été attribuée dans trois matières ou plus (au NS ou au NM).
- Une note finale égale ou inférieure à 3 a été attribuée dans quatre matières ou plus (au NS ou au NM).
- Le candidat a obtenu moins de 12 points pour les matières au NS (seules les trois notes finales les plus élevées comptent pour les candidats inscrits à quatre matières au NS).
- Le candidat a obtenu moins de 9 points pour les matières au NM (les candidats inscrits à deux matières au NM doivent obtenir au moins 5 points au NM).

Le diplôme bilingue de l'IB

À la place du diplôme de l'IB habituel, un « diplôme bilingue » peut être décerné à un candidat qui :

- présente deux langues du groupe 1 et obtient une note finale égale ou supérieure à 3 pour chacune de ces deux langues ;
- présente l'un des cours du groupe 3 ou 4 dans une langue différente de celle choisie pour le groupe 1. Le candidat doit obtenir une note finale égale ou supérieure à 3 à la fois pour la langue du groupe 1 et pour le cours du groupe 3 ou 4.

Les matières pilotes et les matières interdisciplinaires peuvent contribuer à l'obtention d'un diplôme bilingue, si les conditions ci-dessus sont remplies.

Les éléments suivants ne peuvent contribuer à l'obtention d'un diplôme bilingue :

- le mémoire ;
- un programme propre à l'établissement ;

- une matière choisie par le candidat en complément des six matières du diplôme de l'IB (« matière supplémentaire »).

La méthode pour déterminer les notes manquantes

Dans certaines circonstances, il est impossible d'obtenir ou d'utiliser le travail d'un candidat. C'est notamment le cas lorsque le travail a été perdu ou détruit par les services postaux, ou qu'il existe des preuves d'une mauvaise conduite pratiquée à grande échelle sans qu'il soit possible de reconnaître les travaux concernés. Dans de telles situations, l'IB peut tenter d'estimer la note finale qui aurait dû être obtenue, en suivant la procédure pour note manquante.

La procédure générale ainsi que ses points forts et ses points faibles sont expliqués dans la section « [Les notes manquantes](#) ». Le calcul utilisé pour le Programme du diplôme est expliqué ci-après.

Si un établissement compte cinq candidats ou plus inscrits dans la matière concernée, les données de l'établissement sont utilisées pour déterminer les notes manquantes. Si l'établissement a inscrit moins de cinq candidats dans la matière, les données mondiales sont utilisées pour calculer les notes manquantes. L'expérience montre que lorsque l'établissement compte cinq candidats ou plus inscrits dans une matière, le résultat varie généralement peu lorsque l'IB utilise les données de l'établissement ou les données mondiales.

Figure 66

Comment déterminer la note manquante d'un candidat ?

Calcul pour déterminer la note manquante d'un candidat :

$$\text{Quotient} = \frac{\text{Note totale pondérée du candidat pour les autres composantes}}{\text{Note totale pondérée mondiale (ou de l'établissement) pour les autres composantes}}$$

$$\text{Note manquante du candidat} = \text{note pondérée mondiale (ou de l'établissement) pour la composante manquante} \times \text{quotient}$$

Le quotient compare les notes du candidat à la moyenne mondiale ou de l'établissement pour les composantes qu'il a achevées.

- Si les résultats du candidat sont plus élevés que la moyenne, alors ce chiffre sera supérieur à 1.
- Si les résultats du candidat sont plus bas que la moyenne, alors ce chiffre sera inférieur à 1.

Ce mode de calcul, comme toutes les autres procédures pour note manquante, n'est que la « meilleure supposition » possible à partir des preuves disponibles. Il est toujours préférable de pouvoir noter le travail réel d'un candidat et, si nécessaire, de suivre le système de mesure compensatoire.

Le maintien des normes avec des cohortes de taille très variable

Le nombre de candidats inscrits aux matières du Programme du diplôme peut aller de 1 à plus de 40 000. Il ne fait aucun doute que toutes les matières sont d'importance équivalente et que le même usage sera fait des notes finales, qu'un candidat soit le seul à étudier une matière donnée ou qu'il figure au contraire parmi

des milliers. Cependant, pour des raisons pratiques, quelques disparités existent dans les processus s'appliquant aux matières, en fonction du nombre de candidats inscrits.

Le maintien de la qualité du travail des examinateurs

Le modèle d'assurance de la qualité (qui passe par des copies d'entraînement, de validation et de contrôle) vise à garantir que tous les examinateurs ont compris la norme définie par l'examineur principal et qu'ils la suivent. Si les examinateurs sont très peu nombreux, ce modèle peut ne pas être adapté.

Si tous les examinateurs participent à la discussion menée au moment de la normalisation, il est superflu d'utiliser des copies d'entraînement ou de validation, car chaque examinateur aura déjà participé à la définition de la norme. Le recours aux copies de contrôle conserve son utilité pour s'assurer que la notation des examinateurs reste conforme à la norme requise.

Étant donné qu'il pourra être nécessaire de faire appel à des examinateurs supplémentaires si la notation prend trop de temps, un modèle d'assurance de la qualité est généralement prévu même s'il ne s'avère finalement pas utile. Et si aucune copie de validation n'a été préparée, des copies de contrôle pourront être utilisées à la place, même si ce n'est pas idéal, car des éléments différents sont pris en considération lors de la sélection de ces deux types de copies comportant les notes définitives.

Le plus important reste toutefois d'avoir une notation de qualité, même si peu de candidats sont inscrits à une matière, et l'IB se sert de toutes les preuves à sa portée pour détecter les éventuels problèmes avant la publication des résultats.

Les notes finales : le processus complet d'attribution des notes finales (en ligne ou de façon traditionnelle)

Pour les matières qui comptent plusieurs examinateurs en dessous de l'examineur principal dans la hiérarchie, une réunion officielle est organisée pour appliquer les pratiques expliquées dans la section « L'attribution des notes finales (et la totalisation) ». Un responsable de matière de l'IB participe à la réunion pour aider l'examineur en chef et faire un contrôle de la qualité du processus.

Cette réunion peut se dérouler de façon traditionnelle ou en ligne. Dans le second cas, il faut choisir un horaire compatible avec les différents fuseaux horaires associés aux lieux de résidence des examinateurs superviseurs. Ces derniers passent alors par un forum en ligne privé pour échanger leurs points de vue pendant qu'ils étudient les preuves, et par des visioconférences pour prendre les décisions importantes ensemble.

Les notes finales : le processus dirigé d'attribution des notes finales pour un petit nombre de candidats (en ligne)

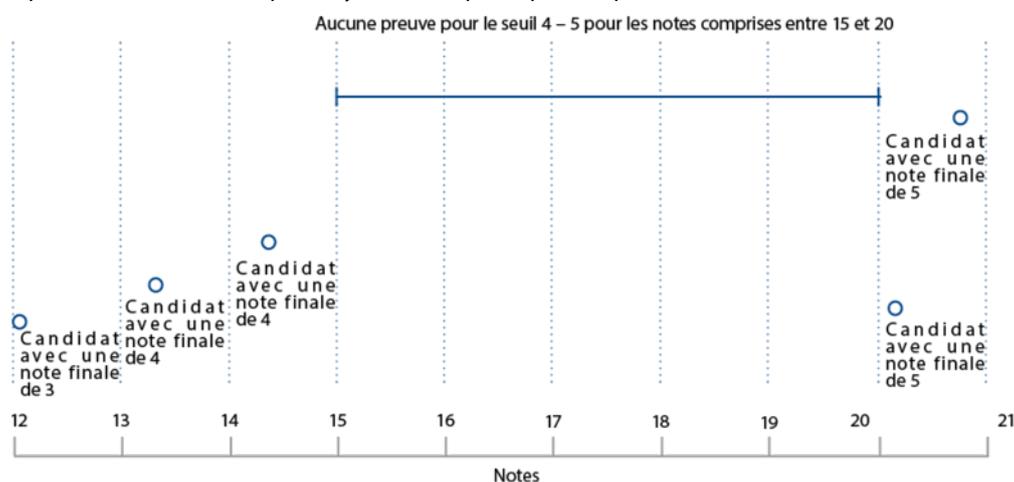
Lorsqu'il existe un nombre très réduit d'examineurs pour chaque composante, les examinateurs superviseurs discutent en ligne. Le responsable de matière de l'IB leur fournit tout l'éventail de preuves nécessaires pour décider convenablement des notes finales à attribuer et contrôle la progression du travail pour veiller au maintien de la qualité, mais il est peu probable qu'il participe aux discussions. En général, les équipes d'examineurs superviseurs chargées de l'attribution des notes finales dans les matières présentées par un petit nombre de candidats comptent également peu de membres.

Les notes finales : le processus classique d'attribution des notes finales pour un petit nombre de candidats

Lorsque le nombre de candidats est trop faible pour que les preuves statistiques accumulées au fil des années soient exploitables, il est demandé aux examinateurs de passer en revue tous les travaux des candidats plutôt que de se concentrer sur un échantillonnage pour recommander les seuils d'attribution des notes finales fondés sur une appréciation. Dans de tels cas, les examinateurs principaux et les examinateurs en chef peuvent suggérer tous les seuils d'attribution des notes finales, et pas uniquement les seuils 2 – 3, 3 – 4 et 6 – 7. Cela s'explique par le manque fréquent de travaux de candidats pour un seuil donné. Ce problème est illustré par l'exemple qui suit.

Figure 67

Exemple de situation dans laquelle il y a un manque de preuves pour un seuil d'attribution des notes finales



Les responsables de matières de l'IB viennent apporter leur soutien à ces réunions si nécessaire, et leur rôle n'est pas d'observer ces réunions mais d'effectuer des contrôles aléatoires de la qualité. Les matières présentées par un petit nombre de candidats sont souvent celles pour lesquelles il est le plus difficile de maintenir une norme constante. En effet, le petit nombre de candidats réduit la valeur des données rétrospectives. De plus, ces candidats se trouvent souvent en haut de la gamme de notes.

Pour faciliter le travail des examinateurs chargés des matières présentées par un petit nombre de candidats (processus dirigé ou classique d'attribution des notes finales), l'IB encourage les discussions entre les examinateurs des différentes matières d'un même groupe. Cela permet de s'assurer d'une compréhension commune de la signification des notes dans tout le groupe de matières.

La validation finale

Quel que soit le nombre de candidats ayant passé les examens, la dernière étape du processus d'attribution des notes finales est la suivante : l'examineur en chef fait des recommandations à l'IB puis les membres de la direction de la division de l'évaluation de l'IB passent en revue les preuves à l'appui de ces recommandations et confirment qu'elles leur conviennent.

L'IB tient compte de l'importance relative des preuves statistiques en fonction du nombre de candidats de la cohorte, mais les matières font toutes l'objet du même examen attentif.

Lectures supplémentaires

Pour obtenir de plus amples informations sur le Programme du diplôme de l'IB, veuillez consulter les ressources suivantes :

- *Le Programme du diplôme : des principes à la pratique ;*
- *Règlement général du Programme du diplôme ;*
- *Procédures d'évaluation du Programme du diplôme ;*
- *Règlement pour les écoles du monde de l'IB : Programme du diplôme ;*
- Les [guides pédagogiques](#) (voir les pages consacrées aux matières sur le Centre de ressources pédagogiques) ;
- Le [matériel de soutien pédagogique](#) publié pour chaque cours (voir les pages consacrées aux matières sur le Centre de ressources pédagogiques).

Le Programme à orientation professionnelle de l'IB

Les caractéristiques de l'évaluation du Programme à orientation professionnelle (POP) sont les suivantes :

- les élèves doivent satisfaire à un certain nombre d'exigences pour obtenir le certificat du POP ;
- aucun résultat général n'est associé au certificat du POP ;
- lorsque des élèves du POP suivent des cours avec des élèves du Programme du diplôme, une évaluation commune a lieu ;
- le cadre pédagogique du POP impose aux élèves de suivre une formation à orientation professionnelle qui n'est pas proposée par l'IB et ne fait pas l'objet d'une sanction officielle des études par l'organisation.

Les objectifs globaux du POP

La validité des résultats de l'évaluation ne peut être déterminée qu'en ayant une idée précise de la finalité des cours et du programme. Cette section commence donc par une présentation des objectifs globaux du programme.

La particularité du POP est qu'il aide les élèves à devenir des apprenants prêts à entrer dans la vie active, et ce, quelle que soit la carrière choisie. Ce programme leur offre la garantie de développer des compétences transférables qui leur seront utiles tout au long de la vie dans l'emploi pour lequel il leur semble bon de se former.

Le POP aide les élèves à :

- développer une vaste gamme de compétences professionnelles et à approfondir leur compréhension dans des domaines spécifiques de la connaissance grâce à leurs cours du Programme du diplôme ;
- développer des stratégies d'acquisition et d'enrichissement de connaissances flexibles dans des contextes variés ;
- se préparer à intégrer efficacement un monde du travail en constante évolution ;
- encourager des attitudes et des habitudes mentales qui leur permettront de devenir de véritables apprenants permanents, désireux d'envisager de nouvelles perspectives ;
- s'impliquer dans un apprentissage qui développe leur capacité et leur volonté d'innover de manière positive.

(Le Programme à orientation professionnelle : des principes à la pratique, 2015, p. 7)

Les utilisations valides des résultats des évaluations du POP

Lorsque l'IB élabore ses modèles d'évaluation et ses programmes d'études, il s'efforce de faire en sorte que le certificat du POP puisse être utilisé pour :

- la sélection en vue de l'obtention d'un emploi et de l'entrée dans un programme professionnalisant comme un apprentissage ;
- la sélection en vue de la poursuite des études dans le domaine professionnel correspondant ;
- la sélection en vue de l'entrée à l'université ;
- déterminer si les candidats ont déjà satisfait aux exigences d'un programme universitaire (attribution de crédits ou dispense de certains modules, cours, etc.).

Lorsqu'un candidat a passé une évaluation dans une langue (d'usage), cela fournit également des preuves de sa capacité à poursuivre ses études dans la matière ou une carrière dans cette langue.

Pour comprendre les raisons pour lesquelles il est important de tenir compte de ces utilisations valides, veuillez vous reporter à la section « Qu'est-ce qui est valide : l'évaluation ou son utilisation ? ».

La structure du POP

Le POP est un cadre pédagogique composé de trois parties, à savoir :

- au moins deux cours du Programme du diplôme, étudiés au niveau moyen (NM) ou au niveau supérieur (NS) ;
- le tronc commun du POP ;
- une formation à orientation professionnelle.

Figure 68

Le modèle du POP : une illustration de la structure du programme



Les composantes du tronc commun, notamment le projet de réflexion

Les composantes du tronc commun visent à contextualiser à la fois les cours du Programme du diplôme et la formation à orientation professionnelle. Le tronc commun se veut le trait d'union entre tous les domaines d'apprentissage.

Il est obligatoire de suivre les quatre composantes, à savoir :

- le cours de compétences personnelles et professionnelles ;
- l'apprentissage par le service ;
- le cours de développement d'une langue seconde ;
- le projet de réflexion.

L'IB évalue uniquement le projet de réflexion, qui est d'abord noté par l'enseignant, puis révisé par l'IB. Pour ce qui est des autres composantes, l'établissement doit confirmer à l'IB que les candidats les ont menées à bien, mais ce dernier ne vérifie aucune évaluation réalisée (le cas échéant).

Les cours du Programme du diplôme dans le cadre du POP

Chaque élève choisit au moins deux matières qui sont également proposées dans le Programme du diplôme. Le même processus d'évaluation s'applique aux élèves du POP et aux élèves du Programme du diplôme. Aucun examen ni aucun processus d'attribution des notes finales n'est spécifique au POP.

Même si les cours sont communs aux deux programmes, un élève ne peut pas s'inscrire à la fois au Programme du diplôme et au POP. La portée des autres exigences spécifiques à chacun des deux programmes exclut toute réalisation simultanée.

La formation à orientation professionnelle

L'IB n'évalue pas les résultats de la formation à orientation professionnelle du POP, et il ne contrôle pas non plus sa qualité. La seule exigence est que l'établissement confirme que l'élève l'a effectivement menée à bien. Les cours du Programme du diplôme de l'IB ne se prêtent pas à une intégration dans la formation à orientation professionnelle du POP.

Le mode de calcul du résultat au POP

Aucun résultat général sous forme de points n'est associé au certificat du POP.

Un candidat se verra décerner le certificat du POP, sous réserve du respect des exigences suivantes :

- l'établissement a confirmé que le candidat a satisfait aux exigences de la formation à orientation professionnelle ;
- le candidat a obtenu une note finale égale ou supérieure à 3 dans au moins deux des cours du Programme du diplôme ;
- le candidat a obtenu une note finale égale ou supérieure à D pour le projet de réflexion ;
- l'établissement a confirmé que le candidat a satisfait à toutes les exigences des autres composantes du tronc commun (cours de compétences personnelles et professionnelles, apprentissage par le service et cours de développement d'une langue seconde) ;
- le candidat n'a pas fait l'objet d'une sanction pour mauvaise conduite de la part du comité d'attribution des notes finales.

Les notes finales des matières du Programme du diplôme et du projet de réflexion sont confirmées par le même comité d'attribution des notes finales que pour le Programme du diplôme.

Le certificat bilingue du POP

À la place du certificat du POP habituel, un « certificat bilingue » peut être décerné à un candidat qui :

- présente deux cours du groupe de matières Études en langue et littérature du Programme du diplôme dans deux langues différentes et obtient une note finale égale ou supérieure à 3 pour chacun de ces deux cours ;
- présente un cours du groupe de matières Études en langue et littérature du Programme du diplôme ainsi qu'un cours du groupe de matières Individus et sociétés ou Sciences du Programme du diplôme dans une langue d'usage différente de celle choisie pour le groupe Études en langue et littérature et obtient une note finale égale ou supérieure à 3 dans ces deux cours.

La gestion des normes et les travaux manquants

Pour le projet de réflexion et tous les cours communs avec le Programme du diplôme, le POP suit les approches du Programme du diplôme, qu'il s'agisse de la procédure pour note manquante ou de la gestion des normes selon la taille de la cohorte.

Lectures supplémentaires

Pour obtenir de plus amples informations sur le Programme à orientation professionnelle de l'IB, veuillez consulter les ressources suivantes :

- *Le Programme à orientation professionnelle : des principes à la pratique ;*
- *Procédures d'évaluation du Programme à orientation professionnelle ;*
- *Aperçu du Programme à orientation professionnelle ;*
- *Guide du projet de réflexion ;*
- *Guide du cours de développement d'une langue seconde ;*
- *Guide du cours de compétences personnelles et professionnelles ;*
- *Guide de l'apprentissage par le service.*

Le Programme d'éducation intermédiaire de l'IB

Les caractéristiques de l'évaluation du Programme d'éducation intermédiaire (PEI) sont les suivantes :

- tous les candidats en 5^e année doivent mener à bien un projet personnel dont la notation est révisée en externe, mais toutes les autres évaluations de l'IB sont facultatives pour les établissements qui proposent le PEI ;
- les candidats doivent étudier un nombre prescrit de matières pour obtenir le certificat du PEI ;
- le résultat global au certificat du PEI est décrit à l'aide d'un total de points sur 56 (maximum) ;
- le projet personnel et l'apprentissage interdisciplinaire (deux composantes du tronc commun) ainsi que les autres disciplines scolaires contribuent à parts égales au résultat global ; le service communautaire (une autre composante du tronc commun) ne contribue pas au résultat global ;
- concernant les évaluations sommatives conçues par l'IB :
- chaque matière ne compte qu'une seule composante,
- les matières sont évaluées au moyen d'un portfolio électronique ou font l'objet d'un examen sur ordinateur.

Les objectifs globaux du PEI

La validité des résultats de l'évaluation ne peut être déterminée qu'en ayant une idée précise de la finalité des cours et du programme. Cette section commence donc par une présentation des objectifs globaux du programme.

Le PEI est conçu comme un cadre pédagogique cohérent et complet qui place les élèves âgés de 11 à 16 ans face à des défis éducatifs et leur permet de développer des compétences importantes dans leur vie d'adolescents. Ces années représentent une période critique dans leur développement, et leur réussite scolaire est liée de près à leur bien-être personnel, social et affectif. À une période où les jeunes cherchent à se construire du point de vue de leur identité et de leur estime de soi, le PEI peut motiver et aider les élèves à réussir dans le contexte scolaire et au-delà. Ce programme leur permet de consolider leurs points forts et de relever des défis dans les matières pouvant leur poser des difficultés. Le PEI donne aux élèves des occasions de développer leur potentiel, de découvrir leurs préférences en matière d'apprentissage et de prendre des risques calculés. Il leur permet également de prendre pleinement conscience de leur identité personnelle et de mener une réflexion sur celle-ci.

(Le Programme d'éducation intermédiaire : des principes à la pratique, 2014, p. 4)

Dans le PEI, les objectifs spécifiques des groupes de matières sont explicitement alignés sur les critères d'évaluation. Tous les groupes de matières du PEI ont quatre critères d'évaluation qui correspondent aux quatre objectifs spécifiques du groupe. Tous les critères contribuent à hauteur équivalente au résultat final.

L'inclusion d'un contexte mondial dans les évaluations électroniques

Dans le cadre du PEI, les contextes d'apprentissage doivent être (ou prendre modèle sur) de véritables événements, situations et circonstances qui surviennent dans le monde. Les contextes d'apprentissage du PEI sont choisis parmi les contextes mondiaux en vue d'encourager la sensibilité internationale et l'engagement mondial au sein du programme. [...] Le PEI identifie six contextes mondiaux pour l'enseignement et l'apprentissage qui s'inspirent des thèmes transdisciplinaires du PP et les renforcent.

(Le Programme d'éducation intermédiaire : des principes à la pratique, 2014, p. 21 et 22)

Chaque session d'examens sera élaborée et orientée en fonction d'un contexte mondial particulier et d'une piste d'exploration sélectionnés dans la liste publiée dans le document intitulé *Le Programme d'éducation intermédiaire : des principes à la pratique*.

Près d'un tiers des tâches de chaque examen sur ordinateur se rapportant à une discipline seront liées au contexte mondial sélectionné, s'en inspireront ou en découleront. La totalité de l'examen sur ordinateur se rapportant à l'apprentissage interdisciplinaire s'inspirera du contexte mondial sélectionné.

Des plans de travail des unités partiellement remplis, qui tiennent compte du contexte mondial sélectionné, seront élaborés pour les groupes de matières Acquisition de langues, Arts, Design, et Éducation physique et à la santé.

Les utilisations valides des résultats des évaluations du PEI

Lorsque l'IB élabore ses modèles d'évaluation et ses programmes d'études, il s'efforce de faire en sorte que les notes finales obtenues dans les cours du PEI et le total des points obtenus au certificat du PEI puissent être utilisés pour :

- la sélection en vue de la poursuite des études ou de l'entrée dans le monde du travail ;
- un retour d'information positif et un résumé des points forts pour les candidats poursuivant leurs études.

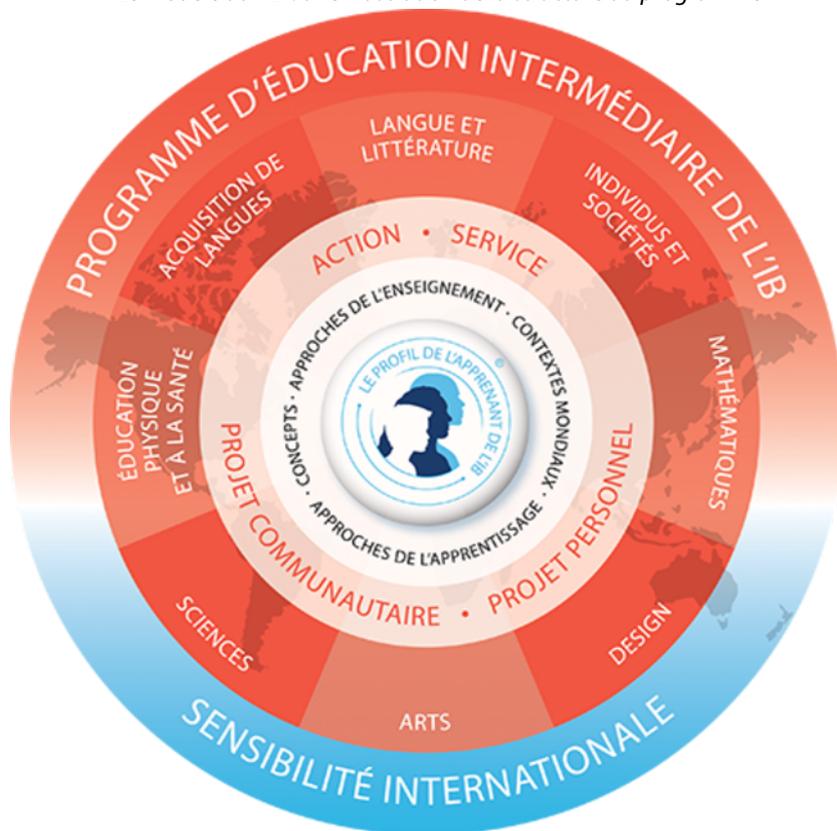
Lorsqu'un candidat a passé une évaluation dans une langue (d'usage), cela fournit des preuves de sa capacité à poursuivre ses études dans cette langue (par exemple, sa capacité à poursuivre ses études dans un établissement d'enseignement français, espagnol ou anglais).

Pour comprendre les raisons pour lesquelles il est important de tenir compte de ces utilisations valides, veuillez vous reporter à la section « [Qu'est-ce qui est valide : l'évaluation ou son utilisation ?](#) ».

La structure du PEI

Figure 69

Le modèle du PEI : une illustration de la structure du programme



Le mode de calcul du résultat au PEI

À la fin de la 5^e année du PEI, les élèves peuvent être inscrits aux évaluations externes de l'IB. Les résultats obtenus aux évaluations seront consignés dans un document présentant les résultats de cours du PEI. Par ailleurs, ces candidats peuvent choisir de passer des évaluations qui conduiront à l'octroi du certificat du PEI.

Les établissements peuvent également délivrer une attestation de participation au PEI. Elle s'adresse aux élèves du PEI qui ont suivi le programme pendant au moins 2 ans et satisfait aux exigences en 3^e ou 4^e année du PEI. Ces élèves ne sont en aucune façon inscrits aux évaluations de l'IB. L'attestation de participation est un document propre à l'établissement qui n'est pas vérifié par l'IB.

Pour pouvoir obtenir le certificat du PEI de l'IB, le candidat doit avoir suivi la dernière année du programme, après une participation recommandée de deux ans, et :

- passer un examen sur ordinateur ou réaliser un portfolio électronique dans six matières choisies dans les groupes suivants : Langue et littérature, Acquisition de langues (ou une deuxième matière du groupe Langue et littérature), Individus et sociétés, Mathématiques, Sciences, et enfin, au choix, Arts, Éducation physique et à la santé ou Design ;
- obtenir une note finale égale ou supérieure à 3 dans chacune des six matières susmentionnées ;
- passer l'examen sur ordinateur pour l'apprentissage interdisciplinaire et obtenir une note finale égale ou supérieure à 3 ;

- achever le projet personnel et obtenir une note finale égale ou supérieure à 3 ;
- obtenir un total général de 28 points ;
- satisfaire aux attentes de l'établissement scolaire concernant le service communautaire.

Pour se voir décerner un certificat bilingue du PEI, le candidat doit en outre obtenir des résultats satisfaisants aux examens sur ordinateur dans l'un des domaines suivants :

- un second cours de Langue et littérature (au lieu d'un cours d'Acquisition de langues) ;
- un cours (ou plus) du groupe de matières Sciences ou Individus et sociétés, ou l'examen interdisciplinaire, dans une langue autre que celle choisie pour le cours de Langue et littérature.

Le déroulement de l'évaluation sommative externe (évaluation électronique du PEI)

L'évaluation électronique facultative du PEI prévoit deux façons d'évaluer les connaissances et les compétences des candidats :

- des portfolios électroniques contenant des travaux réalisés par les candidats dans le cadre des cours d'Acquisition de langues, d'Arts, de Design et d'Éducation physique et à la santé, qui font ensuite l'objet d'une révision de notation pour garantir une norme globale cohérente ;
- des examens sur ordinateur (d'une durée de deux heures chacun) portant sur des cours des groupes de matières Langue et littérature, Individus et sociétés, Sciences et Mathématiques, ainsi que sur l'apprentissage interdisciplinaire.

Par ailleurs, le projet personnel est remis à l'IB par voie électronique et fait l'objet d'une révision de notation. Contrairement aux autres éléments de l'évaluation électronique qui sont facultatifs, le projet personnel doit être réalisé par tous les élèves de 5^e année du PEI.

Les plans détaillés des examens

L'IB publie des plans détaillés des examens pour expliquer précisément aux établissements en quoi consisteront les évaluations électroniques. Ces plans détaillés permettent aux enseignants et aux candidats de comprendre la nature et le but de l'évaluation électronique du PEI. Ils constituent une ressource importante pour aider les candidats à se préparer aux examens sur ordinateur, et à se concentrer sur les critères et les stratégies d'évaluation propres à chaque groupe de matières. Les plans détaillés comportent toujours quatre critères ayant la même pondération.

L'IB s'engage à veiller à ce que pour chaque session, les examens ne s'éloignent pas de plus de trois points de ces plans détaillés.

Les portfolios électroniques et les plans de travail des unités partiellement remplis

Avec les portfolios électroniques, il devient possible d'évaluer les tâches (produits) des travaux réalisés sur une longue période dans le cadre des cours ou des représentations, chose qu'il est par nature difficile d'évaluer lors d'un examen. À la base des portfolios électroniques se trouvent des plans de travail des unités partiellement remplis qui sont suffisamment flexibles pour être adaptés aux contextes locaux et qui aident l'enseignant à s'assurer que le travail du candidat produit des preuves adéquates permettant d'émettre des jugements justes et pertinents. De nouveaux plans de travail des unités partiellement remplis sont fournis chaque session.

Les plans de travail des unités devraient garantir que les tâches décidées par les enseignants permettent aux candidats d'apporter des preuves dans l'ensemble de la gamme de notes finales du PEI. Les évaluations conçues par les enseignants comportent le risque de désavantager les candidats avec un ensemble de tâches anormalement faciles ou difficiles. Lors de la révision de notation, l'IB ne peut attribuer des notes finales que sur la base des travaux de candidats disponibles ; si les tâches conçues par les enseignants correspondent uniquement à une partie de la gamme des descripteurs des notes finales, aucune note finale ne pourra être attribuée en dehors de cette gamme.

L'évaluation unique et la gestion de la charge pesant sur les candidats

Pour les candidats, passer des examens et réaliser des travaux dans le cadre des cours est source de stress et de contraintes. Cela peut également empiéter sur le temps d'enseignement. Pour les candidats de 16 ans qui suivent le PEI, l'IB estime que, tout bien considéré, il est plus adapté de minimiser la part d'évaluation sommative. Cela crée des difficultés, car les candidats n'ont qu'une seule occasion de montrer leurs capacités, mais l'IB les accepte afin de veiller au bien-être général des candidats.

La méthode pour déterminer les notes finales manquantes

L'un des inconvénients de vouloir réduire la charge qui pèse sur les candidats est que, avec l'utilisation d'une seule composante d'évaluation, les preuves disponibles pour le PEI sont très minces. Cela signifie que la procédure pour note finale manquante est moins fiable que pour les autres programmes et donc qu'elle ne doit servir que dans des cas exceptionnels.

- Si un candidat a manqué un examen, il doit pouvoir passer cet examen lors d'une session ultérieure.
- Si un candidat a passé l'examen, mais que des facteurs extérieurs ont pu avoir un effet négatif sur ses résultats, il doit bénéficier de mesures compensatoires.
- Si un candidat s'est conformé aux exigences de l'évaluation et a dûment remis son travail, mais qu'il ne peut pas être noté par l'IB pour des raisons qui échappent au contrôle du candidat ou de l'établissement, l'IB utilise la procédure pour note finale manquante.

Pour que la procédure pour note finale manquante puisse être engagée, le candidat doit avoir obtenu une note finale dans quatre autres cours du PEI au moins. Il est impossible d'estimer des résultats pour des candidats ayant moins de notes finales, car les preuves ne sont pas suffisantes pour pouvoir faire une estimation éclairée.

Pour déterminer la note finale manquante, il convient de calculer la note finale moyenne à partir de toutes les autres matières pour lesquelles il existe une note finale obtenue à partir des travaux réalisés par le candidat au cours des 18 derniers mois.

- Si la décimale est supérieure ou égale à 5, la note finale moyenne doit être arrondie à l'entier supérieur.
- Si la décimale est inférieure à 5, la note finale moyenne doit être arrondie à l'entier inférieur.

Ce mode de calcul, comme toutes les autres procédures pour note manquante, n'est que la « meilleure supposition » possible à partir des preuves disponibles. Il est toujours préférable de pouvoir noter le travail réel d'un candidat et, si nécessaire, de suivre le système de mesure compensatoire.

Lectures supplémentaires

Pour obtenir de plus amples informations sur le Programme d'éducation intermédiaire de l'IB, veuillez consulter les ressources suivantes :

- *Le Programme d'éducation intermédiaire : des principes à la pratique ;*
- *Guide de la session d'examens du PEI ;*
- *les guides pédagogiques du PEI ;*
- *Guide des projets du PEI ;*
- *Procédures d'évaluation du Programme d'éducation intermédiaire ;*
- *Exigences en matière de TI pour l'organisation des examens sur ordinateur ;*
- les activités de familiarisation aux examens sur ordinateur (PC ou Mac).

Le Programme primaire de l'IB

Les caractéristiques de l'évaluation du Programme primaire (PP) sont les suivantes :

- l'évaluation nécessite une collaboration entre les enseignants et les élèves en vue de contrôler l'apprentissage, de le consigner, de le mesurer, d'en rendre compte et de l'adapter ;
- l'IB n'impose aucune évaluation sommative externe.

Les objectifs globaux du PP

La validité des résultats de l'évaluation ne peut être déterminée qu'en ayant une idée précise de la finalité du programme. Cette section commence donc par une présentation des objectifs globaux du programme.

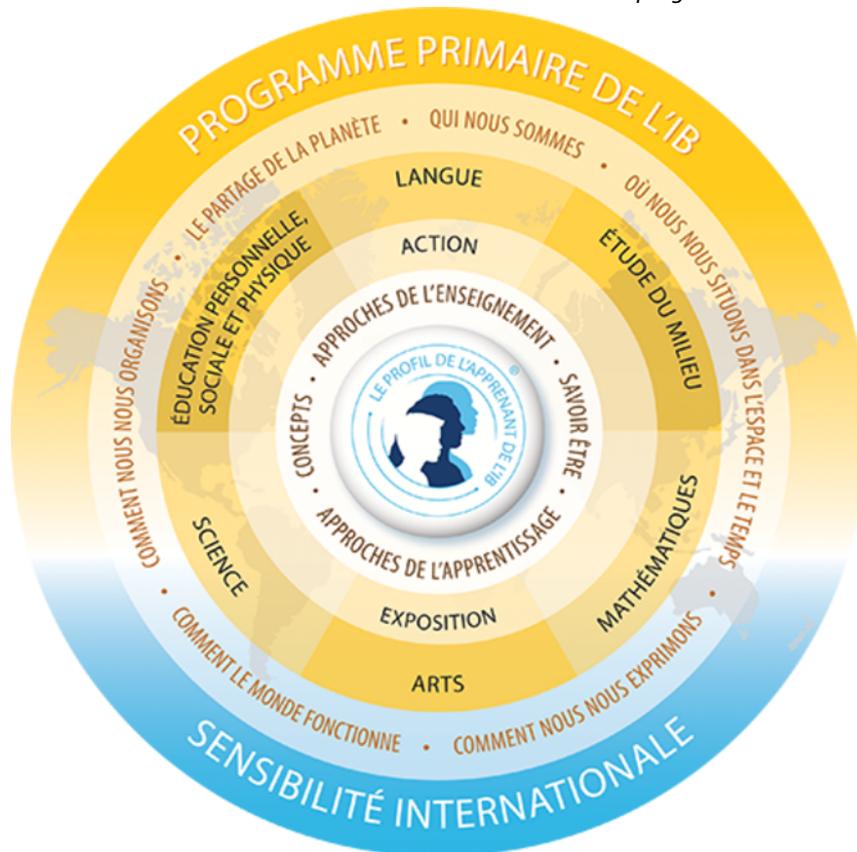
Le PP s'attache aux qualités de cœur autant qu'aux qualités d'esprit et répond aux besoins sociaux, physiques, affectifs et culturels autant qu'aux besoins considérés comme plus scolaires. Les disciplines traditionnelles sont mises en valeur, l'accent étant mis, d'une part, sur l'équilibre entre l'acquisition de connaissances et de savoir-faire essentiels et, d'autre part, sur la quête de sens et de compréhension du monde. Le programme offre la possibilité pour les apprenants de « construire du sens », principalement grâce à une recherche organisée autour de concepts moteurs. Les fils conducteurs de l'apprentissage des élèves sont rassemblés dans le programme de recherche transdisciplinaire qui, à son tour, leur permet d'établir des liens entre ce qu'ils apprennent et la vie hors de l'établissement scolaire.

(Documentation sur le Programme primaire de l'IB)

La structure du PP

Figure 70

Le modèle du PP: une illustration de la structure du programme



L'évaluation formative au service de l'apprentissage

L'objectif premier de l'évaluation au sein du PP est de fournir une rétroaction sur le processus d'apprentissage. Jerome Bruner estime que la rétroaction fournie aux élèves devrait être présentée comme une source d'informations plutôt qu'une récompense ou une punition (Bruner, 1961, p. 26). Les enseignants doivent choisir des stratégies d'évaluation et concevoir des outils d'évaluation qui reflètent précisément les objectifs d'apprentissage dont ils souhaitent rendre compte. Pour évaluer le travail des élèves, ils doivent employer diverses stratégies qui tiennent compte des procédés variés, complexes et recherchés que les élèves utilisent pour comprendre leurs expériences. Enfin, le PP met l'accent sur l'importance de l'autoévaluation et de la réflexion des élèves et des enseignants.

Lectures supplémentaires

Pour obtenir de plus amples informations sur le Programme primaire de l'IB, veuillez consulter les [pages de ressources consacrées au PP](#) sur le Centre de ressources pédagogiques.

Annexe 1 – Révision de notation de l'évaluation interne

La révision de notation est employée pour les travaux évalués en interne afin de veiller à l'application d'une norme commune dans tous les établissements. À l'issue de la révision de notation, les notes de l'établissement peuvent être augmentées ou diminuées, ou elles peuvent rester inchangées. La révision de notation vise à vérifier le degré de précision et de cohérence avec lequel l'enseignant a appliqué les critères d'évaluation lors de la notation des travaux des candidats.

L'échantillonnage

- Au sein de chaque établissement, les enseignants doivent noter les travaux en respectant une norme commune, car, si cela s'avère nécessaire, un facteur de révision de notation peut être appliqué aux notes de tous les candidats étudiant une même matière.
- L'IB tend à ne pas sélectionner les candidats ayant obtenu la note maximale pour l'échantillonnage destiné à la révision de notation. Cela permet aux candidats se situant dans les tranches de notes supérieures d'obtenir une meilleure note à la suite d'une révision de notation.

L'échantillonnage de travaux d'évaluation interne est soigneusement sélectionné de manière à assurer une représentation appropriée de la gamme des notes de l'établissement. Le nombre d'échantillons envoyés pour la révision de notation peut être de dix, huit, cinq ou moins, en fonction du nombre de candidats inscrits dans la matière. L'IB sélectionne les candidats dont les travaux doivent être inclus dans l'échantillonnage après que l'établissement a officiellement envoyé ses notes d'évaluation interne. Comme indiqué précédemment, l'IB tend à ne pas sélectionner les candidats ayant obtenu la note maximale pour la révision de notation afin de permettre aux candidats des tranches de notes supérieures d'obtenir une meilleure note à la suite d'une révision de notation.

Lorsque l'établissement inscrit suffisamment de candidats dans une matière et un niveau donnés pour pouvoir les diviser en plusieurs classes, et que plusieurs enseignants prennent part à l'évaluation interne, l'IB exige de ces enseignants qu'ils se partagent la tâche et travaillent ensemble afin d'uniformiser la façon dont ils se servent des critères d'évaluation. Un seul échantillonnage est alors requis qui, selon toute probabilité, contiendra des travaux notés par les différents enseignants. Toutefois, lorsque les classes utilisent des langues d'usage différentes pour la même matière, un échantillonnage différent doit être soumis pour chaque langue d'usage.

Le calcul des facteurs de révision de notation

Toutes les composantes évaluées en interne sont notées à l'aide de critères d'évaluation ou de bandes de notation et, dans la majorité des cas, les enseignants disposent d'un nombre bien plus important d'informations sur le contexte de travail et la démarche suivie par les candidats pour effectuer ce travail que les examinateurs. De ce fait, les examinateurs qui révisent des composantes évaluées en interne doivent juger du caractère approprié de la notation de l'enseignant et non simplement recorriger le travail sans tenir compte des notes attribuées par l'enseignant. Ces notes ne doivent être modifiées que lorsque les réviseurs sont certains qu'elles sont inappropriées.

L'échantillonnage de travaux notés par l'enseignant est révisé par l'examineur et, sur la base d'une comparaison statistique entre les deux séries de notes (en se servant d'une régression linéaire), un ajustement est apporté aux notes attribuées par l'enseignant (pour cette composante) à tous les candidats de l'établissement.

- Si les notes attribuées par l'enseignant sont systématiquement trop basses ou trop élevées, l'ajustement sera le même pour chacune des notes attribuées par l'enseignant.

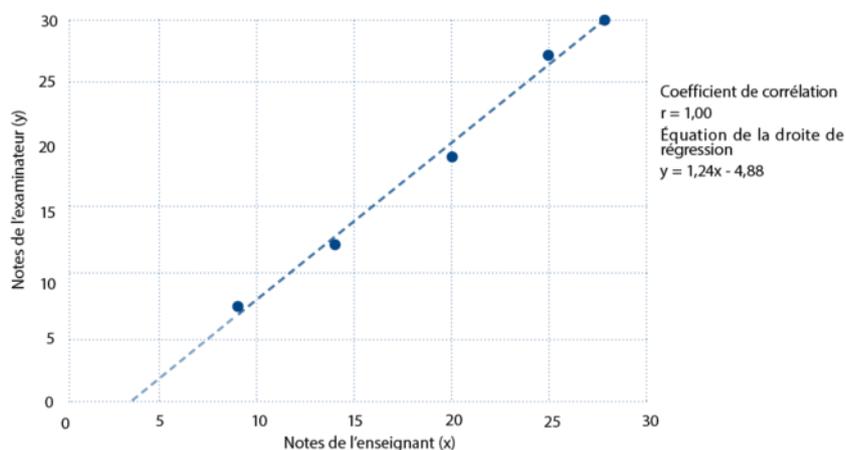
- Si les notes attribuées par l'enseignant sont trop basses ou trop élevées uniquement dans la tranche de notes supérieure ou la tranche de notes inférieure, l'ajustement pourra varier dans toute la gamme de notes.
- Si l'enseignant respecte la norme de notation, aucun ajustement ne sera apporté.

La régression linéaire

Les données recueillies pour chaque échantillonnage envoyé pour la révision de notation font l'objet d'une analyse qui permet d'ajuster de manière appropriée toutes les notes attribuées par l'enseignant, en se basant sur la tendance générale que l'on trouve dans l'échantillonnage. La technique utilisée est appelée « régression linéaire » et consiste à calculer la droite se rapprochant le plus de tous les points du nuage de points de coordonnées qui représente les notes attribuées par l'enseignant et l'examineur pour les travaux inclus dans l'échantillonnage. Un exemple de régression linéaire et l'ensemble de notes correspondant sont présentés ci-dessous.

Figure AN1

Exemple de régression linéaire et ensemble de notes correspondant



Droite de régression de la révision de notation pour un enseignant un peu trop sévère tout en haut et trop généreux en bas. Chaque point représente la paire de notes données à l'un des travaux de l'échantillonnage par l'enseignant et l'examineur. La droite de régression continue est utilisée pour convertir les notes de l'enseignant en notes révisées.

Figure AN2

Les notes correspondantes, issues de la notation de l'enseignant et de celle de l'examineur, sont comparées à l'aide de la régression linéaire. La droite de meilleur ajustement calculée, ou « facteur de révision de notation », donne les notes définitives révisées.

Notes de l'enseignant	Notes de l'examineur	Notes définitives révisées
28	30	30
25	27	26
20	19	20
14	12	12
9	7	6

L'équation de la droite de régression, calculée à l'aide des données de l'échantillonnage, peut être utilisée pour convertir chaque note attribuée par l'enseignant (x) en une note moyenne équivalente (y) que l'examineur aurait très probablement attribuée au même candidat. Cet ajustement apporté à la notation de l'enseignant à la suite de la révision de notation et basé sur l'extrapolation (on passe d'un échantillonnage à un ensemble de notes plus important) peut uniquement refléter les tendances générales observées dans la notation. Les variations liées à un candidat donné ne peuvent pas être prises en considération. Le but de la révision de notation est de s'assurer que les notes des candidats sont, dans l'ensemble, ramenées à des niveaux plus appropriés. La révision de notation ne permet pas de garantir un résultat absolument correct pour chaque candidat.

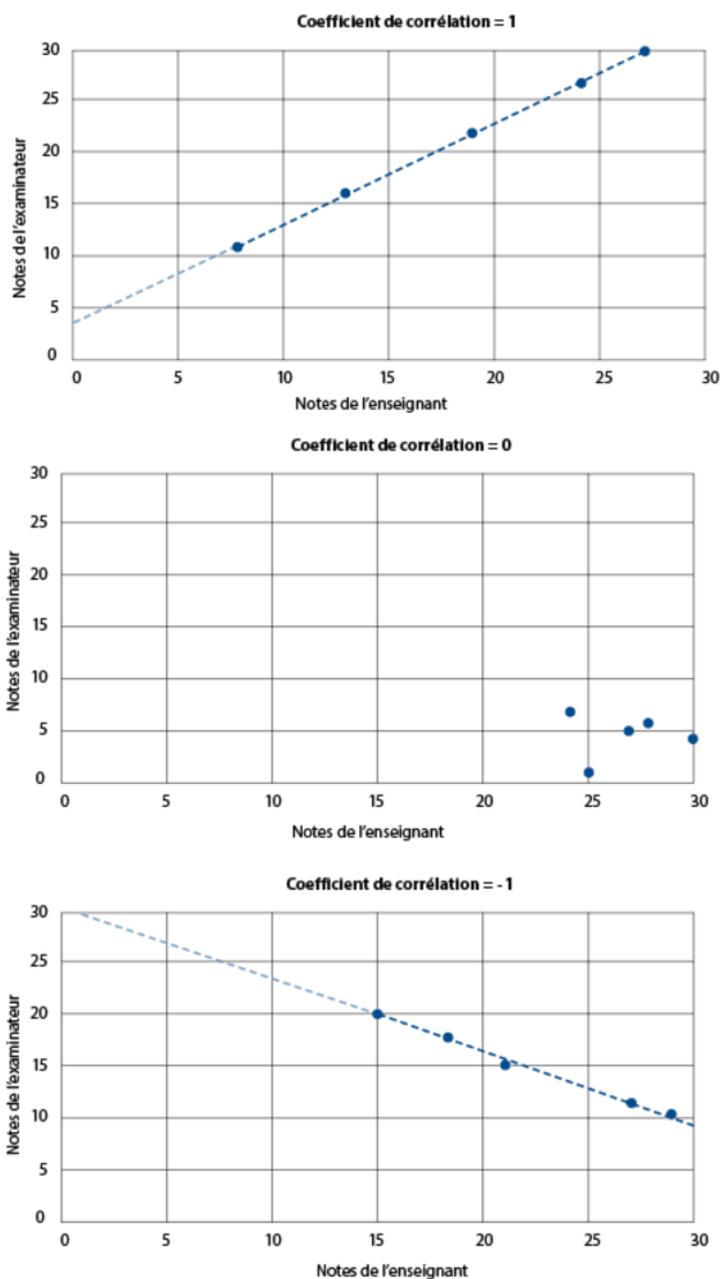
L'échec à la révision de notation

Une vérification est automatiquement effectuée pour contrôler que la droite de régression linéaire (c'est-à-dire le facteur de révision de notation calculé) satisfait à certaines conditions avant d'être appliquée à toutes les notes d'un enseignant. Dans certains cas, il est possible qu'un ajustement de révision de notation ne puisse être calculé sur la base de l'échantillonnage de travaux. L'une des mesures statistiques utilisées est le coefficient de corrélation (coefficient de corrélation de Pearson). Il mesure la cohérence de la relation entre la notation de l'enseignant et celle de l'examineur.

- Un coefficient de corrélation égal à zéro indique qu'il n'y a aucune relation.
- Un coefficient égal à 1 indique une parfaite cohérence dans la relation entre les notes et un accord dans le classement des candidats, du meilleur au plus mauvais (sans pour autant que les notes soient exactement identiques).
- Un coefficient égal à -1 indique un désaccord constant en ce qui concerne le mérite des travaux des candidats, l'enseignant et l'examineur produisant des classements opposés.

Figure AN3

Représentations graphiques du coefficient de corrélation



Pour que le facteur de révision de notation calculé soit acceptable, le coefficient de corrélation doit être égal ou supérieur à 0,85, ce qui indique un haut degré d'accord entre l'enseignant et l'examinateur. Toutefois, un haut coefficient de corrélation seul ne suffit pas à garantir que le facteur de révision de notation est approprié. Il est par ailleurs vérifié que la pente de la droite de régression est comprise entre 0,5 et 1,5. Si la pente de la droite est trop faible, cela signifie que l'enseignant a utilisé une gamme de notes trop large, donnant relativement trop peu de points aux travaux faibles et trop de points aux bons travaux, même si cela est fait de façon cohérente. L'examinateur a donc dû comprimer considérablement la gamme de notes attribuées par l'enseignant. Si la pente est supérieure à 1,5, la droite est trop abrupte, et c'est alors le contraire qui se produit ; l'enseignant n'a pas suffisamment différencié les mauvais et les bons travaux, et l'examinateur a dû élargir la gamme de notes attribuées.

Un échantillonnage « échouera » aux vérifications automatiques de la révision de notation si le coefficient de corrélation est inférieur à 0,85 ou si la pente n'est pas comprise entre 0,5 et 1,5.

Les cas de tous les échantillonnages envoyés par des établissements qui échouent à la révision de notation sont étudiés (au cas par cas) par le personnel de la division de l'évaluation de l'IB, qui examine avec soin les données sous-jacentes et peut décider :

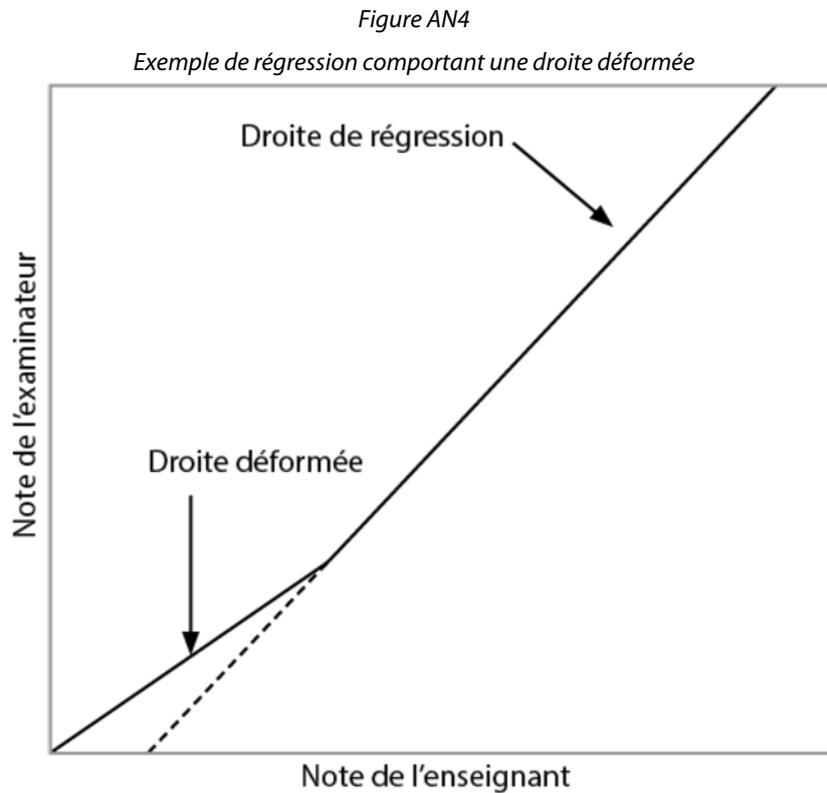
1. que la droite de régression calculée est appropriée pour la gamme de notes particulière de l'enseignant ;
2. d'utiliser une autre sorte d'ajustement pour la révision de notation, qui sera appropriée pour la gamme de notes de l'enseignant ;
3. de demander des données supplémentaires sur l'échantillonnage afin de clarifier la tendance ;
4. de demander les travaux de tous les autres candidats pour procéder à une révision complète de la notation de l'enseignant.

Tous les travaux d'évaluation interne des candidats doivent être disponibles jusqu'à la publication des résultats pour que tous les échecs à la révision de notation puissent être résolus.

L'adaptation du modèle linéaire

Le modèle de la ligne droite suivi pour la révision de notation est, dans une certaine mesure, modifié en utilisant la « déformation ». Il est évident que l'ajustement obtenu grâce à la ligne droite a des effets inappropriés sur les notes extrêmes de la gamme de notes possibles, empêchant un candidat d'obtenir un zéro. Un candidat dont le travail vaut vraiment qu'on lui attribue un zéro peut obtenir quelques points alors que rien dans son travail ne le justifie ou, comme le montre la figure AN4, un candidat ayant produit un travail faible mais valant quelques points peut voir sa note transformée en zéro à la suite de la révision de notation.

Pour résoudre ce problème, on « déforme » la partie extrême de la droite, qui représente les notes situées dans les 20 % inférieurs de la gamme de notes disponibles. À cet extrême, la droite de régression est modifiée pour être remplacée par une nouvelle droite « déformée » qui relie la droite de régression aux points de coordonnées minimum. Cette opération est illustrée dans la figure AN4. La « déformation » de la droite de régression n'est pas effectuée à l'extrémité supérieure de la gamme de notes, car il arrive fréquemment que les enseignants attribuent la note maximum à des candidats qui, de toute évidence, ne le méritent pas.



« Déformation » de la droite de régression pour éviter que les notes des candidats ne soient ajustées en s'éloignant trop ou en s'approchant trop des valeurs minimales.

La « déformation » garantit qu'une note révisée égale à zéro ne peut provenir que d'un zéro attribué par l'enseignant. Cela évite d'attribuer un zéro aux travaux qui valent quelques points et d'attribuer quelques points aux travaux qui n'en valent aucun. Cela suppose bien évidemment que la notation de l'enseignant passe avec succès le processus de révision de notation automatique. Si la notation de l'enseignant « échoue » à la révision de notation et ne peut pas être automatiquement révisée, la droite de régression n'est pas « déformée ».

Annexe 2 – Feuille de route pour créer un argument de validité

L'IB doit être sûr que ses cours et ses évaluations sont adaptés à l'objectif visé. Pour ce faire, il ne peut se contenter de répondre par oui ou non à une série de questions, mais doit engager une discussion sérieuse sur l'équilibre trouvé entre les multiples priorités contradictoires et sur les preuves indiquant que les différents objectifs et buts ont été atteints.

L'« argument de validité » est en réalité un résumé de ces discussions : il rassemble les différents éléments pour expliquer pourquoi l'IB estime que ses cours et ses évaluations sont les meilleurs possible. Les thèmes et les questions ci-dessous constituent le point de départ pour démontrer la validité. Certaines réponses concernent les processus généraux utilisés dans le cadre des évaluations de l'IB, comme les copies de validation et de contrôle, alors que d'autres portent spécifiquement sur certaines matières ou certains cours.

Le processus de création d'un argument de validité n'est pas une fin en soi. Il convient de se poser les bonnes questions et de garder des preuves à chaque étape du cycle, afin que les évaluations et les cours de l'IB soient toujours d'une qualité optimale.

Le développement du cours qui convient le mieux

- Le programme d'études est-il suffisamment approprié et moderne pour permettre une bonne compréhension du cours par les élèves ?
- En quoi ce cours se distingue-t-il des autres cours de l'IB ?
- Comment l'IB sait-il que le programme d'études peut être enseigné efficacement dans les délais impartis ?
- Le processus de révision du programme permet à l'IB de recueillir des preuves pour répondre à ces questions en vue de la formulation de l'argument de validité.

L'alignement des objectifs de l'évaluation sur les objectifs globaux du cours

Que signifie « être bon » dans cette matière ? Comment cette habileté se manifeste-t-elle ?

Qu'est-ce que chaque tâche (composante) est censée évaluer ?

Quel est le degré d'adéquation entre les objectifs globaux du cours et ce qui est évalué ?

En quoi les critères d'évaluation et les descripteurs des notes finales reflètent-ils la manifestation de l'habileté pour chaque tâche ?

Parmi les exemples de preuves possibles, il convient de citer les explications et les justifications du mode d'évaluation, et les notes prises lors des discussions sur les critères et les descripteurs.

L'évaluation du modèle d'évaluation

Comment l'IB sait-il que les normes relatives au programme d'études et à l'évaluation (pas les normes de résultats) sont comparables à celles des autres cours de l'IB ?

En quoi le modèle d'évaluation minimise-t-il les biais ?

Comment les évaluations pourraient-elles être adaptées pour convenir aux candidats ayant des besoins spécifiques (par exemple, des candidats aveugles ou incapables de s'essayer à une tâche donnée) ?

Quel est le niveau de désaccord professionnel acceptable pour la notation ?

Il est possible de trouver des preuves lors de la révision du programme et dans les rapports pédagogiques.

La réflexion sur le processus de préparation des examens (spécimens d'épreuves et épreuves de la session en cours)

En quoi les questions sont-elles un reflet fidèle du programme d'études ?

Parmi les exemples de preuves possibles, il convient de citer la grille d'élaboration, qui décrit la manière dont les questions sont liées au programme d'études et aux objectifs d'évaluation, et les rapports des vérificateurs.

Les dispositions spéciales

Les modalités de l'évaluation la rendent-elles accessible à des candidats aux profils très variés sans qu'ils aient besoin de demander des dispositions spéciales ?

Lorsqu'il a fallu mettre en place des dispositions spéciales, ce besoin a-t-il mis en lumière des problèmes non détectés lors de la phase de conception ?

Comment savoir si les dispositions spéciales mises en place ont permis de maintenir la comparabilité de la norme ?

Parmi les exemples de preuves possibles, il convient de citer les décisions qui ont conduit à l'élaboration des épreuves modifiées ou à l'évaluation de l'accomplissement en adaptant certaines questions des épreuves.

La capacité de gestion des évaluations

Certains problèmes rencontrés par les établissements ou les candidats résultaient-ils des processus de l'IB ou de la conception de l'évaluation ?

Parmi les exemples de preuves possibles, il convient de citer les retours d'information des établissements ou des candidats. Lorsqu'aucun retour d'information n'est envoyé, cela laisse supposer une absence de problème.

L'analyse de la réussite de la notation

Quelle était la fiabilité de la notation ? Les marges de tolérance auraient-elles pu être plus étroites ?

Le barème de notation initial (produit avec l'examen) était-il globalement correct ou a-t-il nécessité des modifications importantes ? Ces modifications étaient-elles raisonnables ou les réponses des élèves auraient-elles dû être prévues lors de la rédaction de l'épreuve (et du barème de notation) ?

Parmi les exemples de preuves possibles, il convient de citer les données sur la validation et le contrôle, et tout remplacement de copies comportant les notes définitives.

La confiance et la comparabilité en matière de décisions concernant l'attribution des notes finales

Dans quelle mesure l'examineur en chef avait-il confiance dans les seuils d'attribution des notes finales recommandés ?

Toutes les preuves soutenaient-elles les décisions prises quant aux seuils d'attribution des notes finales ou différents types d'informations suggéraient-ils des conclusions divergentes ?

Comment l'IB sait-il que les notes finales présentaient une cohérence avec celles des années précédentes et des autres matières ?

Parmi les exemples de preuves possibles, il convient de citer les commentaires contenus dans les rapports internes.

Les preuves de l'équité (biais)

Comment l'IB sait-il que les différents groupes de candidats ont été traités de façon équitable ?

Dans quelle mesure l'IB est-il sûr que la notation et la notation finale ont été cohérentes dans toutes les langues d'usage et zones horaires ?

Parmi les exemples de preuves possibles, il convient de citer l'analyse des données sur les résultats, les commentaires des enseignants, les éléments de preuve apportés par des examinateurs hispanophones et francophones ayant participé à l'attribution des notes finales, les données sur les examinateurs bilingues ayant noté des copies de contrôle en anglais et les données sur l'examen des travaux des candidats de l'autre zone horaire lors de l'établissement des seuils.

Les conséquences des réclamations concernant les résultats

Dans quelle mesure l'IB est-il sûr que les changements apportés aux notes à la suite d'une réclamation concernant les résultats sont raisonnables ?

Dans quelle mesure l'IB est-il sûr que les réclamations concernant les résultats ne montrent aucune erreur systématique dans la notation ou l'attribution des notes finales ?

Parmi les exemples de preuves possibles, il convient de citer l'analyse de tout schéma affiché par les changements apportés à la suite d'une réclamation concernant les résultats ou la mise en évidence d'un même nombre de notes légèrement augmentées et diminuées. L'utilisation des copies de contrôle faite par l'IB durant les recorections découlant de réclamations concernant les résultats sert également de preuve. Des preuves de la participation des examinateurs superviseurs à ce processus pour s'assurer du respect de la norme conviennent également.

La validité prédictive

Quels éléments prouvent que les résultats des élèves sont une bonne indication de leur réussite future dans la matière ? Quels éléments prouvent que cela est une bonne indication de leur réussite dans des matières connexes ?

Parmi les exemples de preuves possibles, il convient de citer les commentaires qualitatifs des élèves et des membres du personnel ainsi que les preuves quantitatives obtenues par l'équipe de recherche de l'IB.

Les preuves d'une culture de l'amélioration continue

Comment l'IB a-t-il tiré des enseignements de la session précédente pour améliorer la session en cours ?

Quelles sont les principales leçons que l'IB peut tirer de la session en cours pour orienter les prochaines sessions ?

Comment l'IB évalue-t-il la réussite de la réalisation de ces améliorations ?

Parmi les exemples de preuves possibles, il convient de citer la façon dont les remarques consignées lors des sessions précédentes ont influencé la conception de la session.

Bibliographie

- ANDERSON, L. W., et KRATHWOHL, D. R. 2001. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York (New York), États-Unis : Longman.
- BAIRD, J., CRESSWELL, M. J., et NEWTON, P. Would the real gold standard please step forward? *Research Papers in Education*. 2000, volume 15, numéro 2, p. 213 – 229.
- BLACK, P. Assessment, Learning Theories and Testing Systems. In MURPHY, P. 1999. *Learners, Learning & Assessment*. Londres, Royaume-Uni : Paul Chapman Publishing en association avec The Open University.
- BLOOM, B. S., ENGELHART, M. D., FURST, E. J., HILL, W. H., et KRATHWOHL, D. R. 1956. *Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain*. New York (New York), États-Unis : Longman.
- BROADFOOT, P. 1996. *Education, Assessment and Society: A Sociological Analysis*. Buckingham, Royaume-Uni : Open University Press.
- BROWN, R. Cultural dimensions of national and international educational assessment. In HAYDEN, M., THOMPSON, J., et WALKER, G. (sous la direction de). 2002. *International education in practice: dimensions for national and international schools*. Londres, Royaume-Uni : Kogan Page.
- CHAMBERLAIN, S. AQA – Public Perceptions of Reliability. 2010. *Ofqual's Reliability Compendium*. Coventry, Royaume-Uni : Office of Qualifications and Examinations Regulation. Chapitre 18.
- COE, R., SEARLE, J., BARMBY, P., JONES, K., et HIGGINS, S. 2008. *Relative difficulty of examinations in different subjects. Report for SCORE (Science Community Supporting Education)*. Durham, Royaume-Uni : CEM Centre, Durham University.
- CRESSWELL, M. J. Examination Grade: How many should there be? *British Educational Research Journal*. 1986, volume 12, numéro 1.
- CRESSWELL, M. J. Defining, setting and maintaining standards in curriculum-embedded examinations: Judgemental and statistical approaches. In GOLDSTEIN, H., et LEWIS, T. (sous la direction de). 1996. *Assessment: Problems, Developments and Statistical Issues*. Chichester, Royaume-Uni et New York, États-Unis : Wiley. P. 57 – 84.
- CRESSWELL, M. J. The role of public examinations in defining and monitoring standards. *Proceedings of the British Academy*. 2000, volume 102, p. 69 – 120.
- CROOKS, T. J., KANE, M. T., et COHEN, A. S. Threats to the Valid Use of Assessments. *Assessment in Education: Principles, Policy and Practice*. 1996, volume 3, numéro 3.
- DOLAN, R. P., BURLING, K., HARMS, M., STRAIN-SEYMOUR, E., WAY, W., et ROSE, D. H. 2013. *A Universal Design for Learning-based framework for designing accessible technology-enhanced assessments*. Compte-rendu de la recherche. Iowa City (Iowa), États-Unis : Pearson Education Measurement.
- FRITH, D. S., et MACINTOSH, H. G. 1984. *A Teachers' Guide to Assessment*. Cheltenham, Royaume-Uni : Stanley Thomas.
- GIBBS, G. 1992. *Assessing More Students*. Oxford, Royaume-Uni : Oxford Centre for Staff Learning and Development.
- GIPPS, C., et MURPHY, P. 1994. *A Fair Test? Assessment, Achievement and Equity*. Buckingham, Royaume-Uni : Open University Press.
- GLASER, R. Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*. 1963, volume 18, p. 519 – 521.
- GOLDSTEIN, H. Group differences and bias in assessment. In GOLDSTEIN, H., et LEWIS, T. (sous la direction de). 1996. *Assessment: Problems, Developments and Statistical Issues*. Chichester, Royaume-Uni et New York, États-Unis : Wiley. P. 85 – 93.

- GOOD, F.J., et CRESSWELL, M.J. 1988. *Grading the GCSE*. Londres, Royaume-Uni : Secondary Examinations Council.
- HE, Q., OPPOSS, D., et BOYLE, A. 2010. A Quantitative Investigation into Public Perceptions of Reliability in Examination Results in England. *Ofqual's Reliability Compendium*. Coventry, Royaume-Uni : Office of Qualifications and Examinations Regulation. Chapitre 19, p. 68.
- HIERONYMOUS, A.N., et HOOVER, H.D. 1986 *Iowa tests of basic skills manual for school administrators*. Iowa City, États-Unis : University of Iowa.
- HUGHES, D., KEELING, B., et TUCK, B. Effects of achievement expectations and handwriting quality on scoring essays. *Journal of Educational Measurement*. 1983, volume 20, numéro 1, p. 65 – 70.
- HUMPHREYS, L. G. An analysis and evaluation of test and item bias in the prediction context. *Journal of Applied Psychology*. 1986, volume 71, p. 327 – 333.
- LAMBERT, D., et LINES, D. 2000. *Understanding Assessment*. Londres, Royaume-Uni : Routledge Falmer.
- LINN, M. C. Gender differences in educational achievement. In PFLEIDERER, J. (sous la direction de). 1992. *Sex Equity in Educational Opportunity, Achievement, and Testing*. Princeton (New Jersey), États-Unis : Educational Testing Service.
- LLEWELLYN, D. *et al.* 2014. *Inquire within: Implementing inquiry and argument-based science standards in grades 3 – 8*. 3^e édition. Thousand Oaks (Californie), États-Unis : Corwin.
- MEYER, A., ROSE, D. H., et GORDON, D. 2014. *Universal design for learning: Theory and Practice*. Wakefield (Massachusetts), États-Unis : CAST Professional Publishing.
- MEIGHAN, R. *Restructuring Education so it works for kids & society* [en ligne]. Avril 2014. Référence du 15 janvier 2017. Disponible sur Internet : <<http://lifelearningmagazine.com/0412/restructuring-education.htm>>.
- MITRA, S. 13 – 16 octobre 2011. Discours prononcé lors de la conférence des directeurs d'école du monde de l'IB à Singapour.
- MURPHY, P. 1999. *Learners, Learning & Assessment*. Londres, Royaume-Uni : Paul Chapman Publishing en association avec The Open University.
- National Research Council. 2011. *Assessing 21st Century Skills: Summary of a Workshop*. Committee on the Assessment of 21st Century Skills. Washington D.C., États-Unis : The National Academies Press.
- NEWTON, P.E. Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy and Practice*. 2007, volume 14, numéro 2, p. 149 – 170.
- NEWTON, P. E. 2012. We need to talk about Validity. Communication présentée lors de la réunion annuelle du National Council for Measurement in Education à Vancouver, Canada.
- NUTTALL, D. L., BACKHOUSE, J. K., et WILLMOTT, A. S. 1974. Comparability of standards between subjects: Schools Council Examinations Bulletin 29. Londres, Royaume-Uni : Evans/Methuen Educational.
- OCDE. *21st century skills and competences in OECD countries* [en ligne]. Paris, France : OCDE, décembre 2009. Référence du 31 janvier 2017. Disponible sur Internet : <<https://www.oecd.org/edu/ceri/44254873.ppt>>.
- OFQUAL. 2015. *Inter-Subject Comparability: A Review of the Technical Literature: ISC Working Paper 2*. Coventry, Royaume-Uni : Office of Qualifications and Examinations Regulation.
- PETERSON, A.D.C. 1971. *New Techniques for the Assessment of Pupils' Work*. Strasbourg, France : Conseil de l'Europe. [Cet ouvrage est disponible en français sous le titre *Nouvelles techniques d'évaluation du travail scolaire*.]
- PETERSON, A.D.C. 2003. *Schools Across Frontiers: The Story of the International Baccalaureate and the United World Colleges* (2^e édition). Chicago, États-Unis : Open Court Publishing Company.
- POPHAM, W. J. 1978. *Criterion-referenced Measurement*. Englewood Cliffs (New Jersey), États-Unis : Prentice Hall.
- RAND EDUCATION. *Teaching and Learning 21st Century Skills: Lessons from the Learning Sciences* [en ligne]. New York (New York), États-Unis : Asia Society, avril 2012. Référence du 1^{er} janvier 2017. Disponible sur Internet : <http://www.rand.org/pubs/external_publications/EP51105.html>.

- RAO, K., CURRIE-RUBIN, R., et LOGLI, C. 2016. *UDL and Inclusive Practices in IB Schools Worldwide*. Wakefield, États-Unis : CAST Professional Learning.
- SCHLEICHER, A. *The case for 21st-century learning* [en ligne]. OCDE, 2016. Référence du 1^{er} janvier 2017. Disponible sur Internet : <http://www.rand.org/pubs/external_publications/EP51105.html>.
- SHEPARD, L. A. Commentary: what policy makers who mandate tests should know about the new psychology of intellectual ability and learning. In GIFFORD, B. R. et O'CONNOR, M. C. (sous la direction de). 1992. *Changing Assessments: Alternative Views of Aptitude, Achievement and Instruction*. Boston, États-Unis et Dordrecht, Pays-Bas : Kluwer. P. 301 – 328.
- SIRECI, S. G. On test validity theory and test validation. *Educational Researcher*. 2007, volume 36, numéro 8, p. 477 – 481.
- SMITH, D., et TOMLINSON, S. 1989. *The School Effect*. Londres, Royaume-Uni : Policy Studies Institute.
- SNYDER, B. R. 1971. *The Hidden Curriculum*. Cambridge (Massachusetts), États-Unis : MIT Press.
- SURGENOR, P. 2010. *Teaching Toolkit: Effect of Assessment on Learning*. Dublin, Irlande : UCD Dublin.
- VYGOTSKY, L.S. 1962. *Thought and Language*. New York (New York), États-Unis : Wiley.
- WILIAM, D. Validity, dependability and reliability in National Curriculum assessment. *The Curriculum Journal*. 1993, volume 4, numéro 3, p. 335 – 350.
- WINKLEY, J., et CRESSWELL, M. 2010. Introduction to the concept of Reliability. 2010. *Ofqual's Reliability Compendium*. Coventry, Royaume-Uni : Office of Qualifications and Examinations Regulation. Chapitre 1.
- WOOD, D., BRUNER, J. S., et ROSS, G. The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*. 1976, volume 17, p. 89 – 100.
- WOOD, D. 1998. *How Children Think and Learn: The Social Contexts of Cognitive Development*. 2^e édition. Oxford, Royaume-Uni : Blackwell.
- WOOD, R. 1991. *Assessment and Testing: A survey of research*. Cambridge, Royaume-Uni : Cambridge University Press.

Glossaire

Terme	Définition
Alignement	Accord de principe qui s'opère également dans la pratique entre des valeurs et idéaux en matière d'apprentissage (le programme établi), l'enseignement effectif (le programme enseigné) et l'apprentissage réalisé par les élèves (le programme évalué).
Aménagements à des fins d'accès et d'inclusion	Conditions modifiées ou supplémentaires lors de la procédure d'évaluation d'un candidat ayant des besoins en matière d'aménagement de la procédure d'évaluation. Ces aménagements permettent au candidat de montrer son niveau d'accomplissement de manière plus équitable. Ils ne sont pas destinés à compenser le moindre manque d'aptitude.
Approche critériée	Comparaison de l'accomplissement du candidat avec des descriptions prédéfinies de l'accomplissement (critères) pour la notation.
Approche critériée limitée	Approche consistant à comparer l'accomplissement du candidat à des descriptions prédéfinies de l'accomplissement (critères) et à l'accomplissement des candidats des cohortes précédentes. C'est l'approche adoptée par l'IB pour garantir le maintien des normes.
Approche normative	Approche dans laquelle le résultat est déterminé en comparant l'accomplissement du candidat à celui de l'ensemble de la population pour laquelle l'évaluation est conçue. Argument de validité Preuve et explication des décisions prises lors de la création d'une évaluation, qui justifient son adéquation avec l'objectif visé.
Attribution des notes finales	Le processus d'attribution des notes finales détermine comment convertir les notes en notes finales pour garantir que ces dernières ont la même signification, quelle que soit la session pendant laquelle un candidat passe ses examens.
Auteur d'épreuve	Personne élaborant les questions et le barème de notation associé, qui seront utilisés pour les évaluations.
Authentification	Processus prouvant que le travail a été produit par le candidat. Il peut s'agir, par exemple, de la signature de l'enseignant et du candidat pour attester la provenance de la réponse. Bande de notation Gamme (prescrite ou donnée) de notes qui doivent

Terme	Définition
	être attribuées à une réponse laissant apparaître certaines qualités.
Barème de notation	Indication de la façon d’attribuer les niveaux des critères pour une tâche particulière. Barème de notation analytique Barème de notation indiquant les bonnes réponses et les éléments pour lesquels il faut attribuer des points.
Besoins en matière d’aménagement de la procédure d’évaluation	Un candidat ayant des besoins en matière d’aménagement de la procédure d’évaluation requiert une modification des conditions d’évaluation afin de pouvoir montrer son niveau d’accomplissement.
Biais	Le biais se manifeste lorsqu’un groupe défini (par exemple, un groupe ethnique ou racial, ou un groupe de personnes du même sexe) obtient des résultats différents de ceux de la moyenne à une question ou une tâche particulière pour une raison autre que sa compétence dans la caractéristique évaluée.
Candidat	Élève inscrit à une évaluation.
Capacité de gestion	Mesure dans laquelle l’évaluation et chaque tâche constituent une charge pour l’élève ou l’établissement. La longueur des évaluations, l’équipement ou le matériel requis pour réaliser l’évaluation ou le nombre d’évaluations requises dans un programme sont autant d’exemples de facteurs devant être pris en considération pour déterminer la capacité de gestion.
Chef d’équipe	Examineur qui dirige une équipe d’examineurs.
Circonstances défavorables	<p>Les circonstances défavorables sont celles échappant au contrôle du candidat et susceptibles d’être préjudiciables à ses résultats à l’issue de l’évaluation. Ce sont, par exemple, un stress grave, des circonstances familiales particulièrement éprouvantes, un deuil ou des événements pouvant menacer la santé ou la sécurité du candidat. Il arrive que des circonstances défavorables concernent un groupe ou l’ensemble des candidats dans un établissement scolaire donné. Les circonstances défavorables n’incluent pas :</p> <ul style="list-style-type: none"> • les insuffisances dues au fait de l’établissement scolaire où le candidat est inscrit ; • l’incapacité des candidats à améliorer leurs résultats malgré les aménagements à des fins d’accès et d’inclusion autorisés dont ils ont bénéficié.
Circonstances exceptionnelles	Circonstances qui ne font pas partie de l’expérience courante des autres candidats ayant des besoins en

Terme	Définition
	matière d'aménagement de la procédure d'évaluation. L'IB se réserve le droit de déterminer les circonstances à qualifier « d'exceptionnelles » qui justifieraient par conséquent un aménagement particulier à des fins d'accès et d'inclusion.
Comparabilité	Mesure dans laquelle un résultat particulier peut être considéré comme identique à un autre résultat. Elle permet généralement de comparer deux années (mesure dans laquelle la note finale 8 est comparable à la note finale 8 de l'année précédente) ou deux matières (mesure dans laquelle la note finale 5 en mathématiques est comparable à la note finale 5 en arts).
Composante	Voir Composante d'évaluation .
Composante d'évaluation	Une composante d'évaluation est constituée d'une ou plusieurs tâches groupées ensemble, qui représentent une partie de l'évaluation globale. Il peut s'agir, par exemple, d'une épreuve d'examen, d'un dossier de travail, d'un projet ou d'un travail de recherche. Comprendre la validité Terme générique décrivant la pertinence d'une évaluation ou du but pour lequel les résultats de l'évaluation sont utilisés.
Comprendre la validité	Terme générique décrivant la pertinence d'une évaluation ou du but pour lequel les résultats de l'évaluation sont utilisés.
Conception universelle de l'évaluation	Concept selon lequel toutes les évaluations devraient être conçues sur la base d'une compréhension des différents besoins des candidats au lieu de traiter certains candidats différemment. Cette idée fait partie intégrante de l'engagement de l'IB envers la conception universelle de l'apprentissage (CUA).
Copie	Réponses du candidat aux questions de l'épreuve d'examen. Il peut également s'agir d'un travail envoyé par le candidat pour évaluation.
Copie d'entraînement	Exemple de travail d'un candidat choisi et noté pendant la normalisation, puis fourni aux examinateurs afin de leur expliquer la norme de notation qu'ils doivent adopter.
Copie de contrôle	Une copie de contrôle est une copie d'examen qui a déjà été notée par l'examineur principal et qui est incluse de manière aléatoire dans un paquet de copies assignées à un examinateur pour notation. Elle ressemble à n'importe quelle autre copie et l'examineur ne peut donc pas savoir qu'il s'agit d'une copie de contrôle. Les points attribués à cette copie de contrôle par l'examineur seront comparés à ceux attribués par l'examineur principal, en laissant une certaine marge de

Terme	Définition
	tolérance, afin de vérifier que l'examineur respecte la norme de notation établie. Dans le cadre du processus de révision de notation, des copies de contrôle seront incluses de la même façon dans l'échantillonnage dynamique.
Copie de validation	Exemple de travail d'un candidat sélectionné par l'examineur principal qui est utilisé pour vérifier de manière officielle que les examinateurs ont compris la norme de notation à appliquer avant d'être autorisés à noter les copies d'examen de la session en cours.
Copie électronique	Réponses du candidat à une évaluation électronique.
Cours	Nombre prescrit de leçons ou d'heures d'enseignement au sein d'une période d'enseignement déterminée. Les établissements organisent l'enseignement et l'apprentissage dans les matières au moyen de cours disciplinaires et interdisciplinaires.
Critères d'évaluation	Critères permettant de mesurer l'accomplissement d'un candidat. Critères globaux Approche adoptée pour évaluer le travail d'un candidat, qui consiste à examiner le travail dans son ensemble au lieu d'examiner séparément des éléments distincts (par exemple, communication, connaissances disciplinaires, qualité de l'argumentation, etc.).
Critères globaux	Approche adoptée pour évaluer le travail d'un candidat, qui consiste à examiner le travail dans son ensemble au lieu d'examiner séparément des éléments distincts (par exemple, communication, connaissances disciplinaires, qualité de l'argumentation, etc.).
Cycle d'évaluation	Mesures prises pour créer, effectuer et noter les évaluations, y compris les examens. Il s'agit d'un cycle, car l'IB tire des enseignements de chaque session d'examens afin d'améliorer les sessions suivantes.
Descripteurs de niveaux	Les descripteurs de niveaux décrivent les caractéristiques devant être présentes dans le travail du candidat pour chaque niveau.
Descripteurs des notes finales	Ensemble des qualités attendues des candidats pour chaque note finale. Les descripteurs des notes finales peuvent être spécifiques à une matière, spécifiques à un groupe de matières ou généraux pour s'appliquer à l'ensemble du programme. Dans tous les cas, ils doivent décrire les mêmes caractéristiques. Les exemples plus précis ne font qu'expliquer ce que ces descriptions signifient dans le contexte spécifique à la matière.

Terme	Définition
Différenciation (enseignement et apprentissage)	Modification des stratégies d'enseignement pour répondre aux divers besoins d'apprentissage des élèves, en utilisant des contenus, processus et outils variés. Différenciation (évaluation) Action de distinguer des candidats montrant différents niveaux de compétence.
Différenciation (évaluation)	Action de distinguer des candidats montrant différents niveaux de compétence.
Discipline	Domaine d'apprentissage ou champ d'études scolaires. Manière d'organiser les connaissances à des fins d'enseignement (pour des raisons pratiques liées à l'évaluation, les disciplines sont généralement appelées « matières » dans le contexte du PEI et du Programme du diplôme). Certains groupes de matières et certaines matières du PEI sont composés de plusieurs disciplines. Par exemple, le groupe de matières Arts du PEI intègre des disciplines telles que les arts visuels, le théâtre, la musique, les médias et la danse. La matière intitulée « cours intégré de sciences » regroupe quant à elle trois disciplines, à savoir la biologie, la chimie et la physique.
Échantillonnage dynamique	Une amélioration du processus de révision de notation permettant une meilleure utilisation des contrôles de qualité. Il permet d'appliquer la « marge de tolérance » du modèle d'assurance de la qualité aux notes des enseignants et des examinateurs. En ce qui concerne les enseignants, si la notation de l'échantillonnage initial se situe dans la marge de tolérance, aucun facteur de révision de notation n'est appliqué. Cela signifie également que les réviseurs de notation (examineurs) reçoivent des travaux de candidats distincts qui permettent d'inclure des copies de contrôle afin de maintenir une norme uniforme. Cela permet aussi aux examinateurs de recevoir le nombre nécessaire de copies supplémentaires s'il s'avère que l'enseignant ne respecte pas la norme globale.
Échantillonnage envoyé pour la révision de notation	Échantillonnage de travaux de candidats qui est envoyé à l'IB pour s'assurer que les travaux sont notés conformément à la norme attendue.
Effet de remous	Impact que les parties ultérieures d'un processus ont sur les parties antérieures. Dans le contexte de l'éducation, l'effet de remous désigne généralement la façon dont l'enseignement et l'apprentissage sont modifiés par la manière dont le candidat est évalué.
Élément	La plus petite unité d'une tâche ou question d'évaluation. Chaque élément se voit attribuer un certain nombre de points. Un élément peut prendre

Terme	Définition
	la forme d'une question dans son ensemble ou de plusieurs parties d'une question.
Élève	Personne suivant un cours ou un programme de l'IB. Les élèves deviennent des candidats lorsqu'ils prennent part au processus d'évaluation.
Épreuve	Voir Épreuve d'examen .
Épreuve d'examen	Ensemble de tâches et de questions que le candidat doit réaliser et traiter. Dans certains cas, ce terme peut désigner un examen sur ordinateur ou un examen sur papier.
Épreuve modifiée	Modifications apportées à une évaluation pour permettre à un candidat ayant des besoins particuliers de réaliser cette évaluation sur un pied d'égalité avec les autres candidats n'ayant pas les mêmes besoins. Il peut s'agir, par exemple, de modifier la taille ou le style de la police de caractères. Ces ajustements ne doivent pas modifier la nature de la question posée.
Évaluation	Recueil de preuves afin d'émettre des jugements sur l'enseignement et l'apprentissage. Évaluation critériée Processus d'évaluation selon lequel les niveaux sont déterminés à l'aide de critères précédemment convenus. La norme de notation est donc fixe et ne dépend pas des accomplissements de la cohorte dans son intégralité.
Évaluation critériée	Processus d'évaluation selon lequel les niveaux sont déterminés à l'aide de critères précédemment convenus. La norme de notation est donc fixe et ne dépend pas des accomplissements de la cohorte dans son intégralité.
Évaluation électronique	Évaluation effectuée au moyen d'un ordinateur ou d'un appareil similaire. Évaluation externe Évaluation organisée et notée par l'IB, et non par l'enseignant du candidat.
Évaluation externe	Évaluation organisée et notée par l'IB, et non par l'enseignant du candidat.
Évaluation finale	Évaluation sommative du travail des élèves à la fin de la période d'enseignement.
Évaluation formative	Évaluation continue destinée à fournir des informations qui orienteront l'enseignement et permettront à l'élève d'améliorer son travail.
Évaluation inclusive	Évaluation tenant compte des besoins de tous les candidats, afin qu'ils puissent montrer de manière équitable leurs compétences dans la matière.
Évaluation interdisciplinaire	Combinaison ou regroupement d'au moins deux domaines d'apprentissage ou champs d'études scolaires dans une même évaluation. Dans le

Terme	Définition
	Programme du diplôme, une matière interdisciplinaire se conforme aux exigences de deux groupes par l'entremise d'une seule matière. Dans le cadre du PEI, l'étude interdisciplinaire peut être développée au sein des groupes de matières et entre eux. L'évaluation externe interdisciplinaire du PEI fait toujours intervenir plusieurs groupes de matières.
Évaluation interne	Évaluation effectuée par les enseignants dans l'établissement. Évaluation sommative Évaluation visant à déterminer la compétence ou le niveau atteint par les élèves, qui est généralement effectuée à la fin du programme d'études ou d'une unité de travail.
Évalué en externe	Travail évalué ou noté entièrement par l'IB. Évalué en interne Travail qui est évalué (noté) par l'enseignant des élèves. L'IB demande aux établissements d'envoyer un échantillonnage de travaux évalués en interne pour la révision de notation.
Évalué en interne	Travail qui est évalué (noté) par l'enseignant des élèves. L'IB demande aux établissements d'envoyer un échantillonnage de travaux évalués en interne pour la révision de notation.
Examen	Une ou plusieurs tâches de types divers (questions à réponse brève, questions à réponse développée, exercices de résolution de problèmes ou questions d'analyse ; parfois, exercices pratiques ou tâches orales) que les candidats doivent effectuer seuls, et ce, en un temps donné et sous surveillance.
Examen sur ordinateur	Examen officiel, en temps limité, élaboré en externe et s'appuyant sur de nombreux médias, qui comprend une série de tâches en rapport avec la matière. Ces tâches sont effectuées dans un environnement d'examen protégé. Examineur Personne qui attribue des points au travail produit par le candidat pour l'évaluation externe.
Examineur	Personne qui attribue des points au travail produit par le candidat pour l'évaluation externe.
Examineur en chef	Examineur le plus haut placé dans la hiérarchie, qui est chargé de veiller au respect des normes de notation au fil du temps et entre les disciplines d'un groupe de matières (par exemple, Sciences).
Examineur principal	L'examineur principal est chargé de diriger l'évaluation dans une composante. Il établit la norme d'évaluation et, en général, il est aussi l'un des auteurs de l'évaluation.

Terme	Définition
Examineur superviseur	Fonction des examinateurs expérimentés qui aident l'examineur principal.
Facteur de révision de notation	Coefficient mathématique appliqué aux totaux des niveaux par critère d'un évaluateur pour les rendre conformes à la norme d'évaluation commune.
Fiabilité	Mesure dans laquelle un candidat se verra attribuer la même note chaque fois que son travail sera évalué. Il peut s'agir de la fiabilité de la notation des examinateurs (c'est-à-dire la mesure dans laquelle ils donnent tous la même note au candidat) ou de la fiabilité de la notation d'un seul examinateur (c'est-à-dire la mesure dans laquelle il donne la même note au candidat chaque fois qu'il corrige son travail).
Fraude	Toute pratique contrevenant aux principes de l'intégrité intellectuelle (par exemple, le plagiat).
Groupe d'éléments de question (QIG pour <i>Question Item Group</i> en anglais)	Une ou plusieurs questions connexes au sein d'une épreuve d'examen considérées comme un groupe pour la notation. Les examinateurs notent ce groupe d'éléments de question plutôt que l'épreuve dans son intégralité. Cette approche permet une notation plus fiable que les approches qui consistent à noter la copie dans son intégralité.
IBIS	Site Web permettant aux coordonnateurs des établissements d'accomplir des tâches administratives et d'obtenir des informations de l'IB par l'intermédiaire d'un serveur protégé par mot de passe.
Inscription des candidats	Procédure entreprise par le coordonnateur de l'établissement pour inscrire les candidats à une évaluation de l'IB.
Intégrité intellectuelle	L'intégrité intellectuelle est un ensemble de valeurs et de compétences, qui encourage l'intégrité personnelle et les bonnes pratiques en matière d'enseignement, d'apprentissage et d'évaluation.
Jugement	Comparaison faite entre le travail d'un candidat et un critère d'évaluation.
Langue d'usage	Langue dans laquelle le candidat effectue son évaluation.
Langues de travail	Langues utilisées par l'IB pour communiquer avec ses interlocuteurs et dans lesquelles elle s'engage à fournir toute une gamme de services pour la mise en œuvre de ses programmes. Ces langues sont actuellement l'anglais, le français et l'espagnol.
Liste de questions	Ensemble de questions accompagnées d'informations sur le thème et le degré de difficulté attendu. Les informations fournies dans une liste de questions peuvent être utilisées pour créer les

Terme	Définition
	épreuves d'examen. À l'heure actuelle, l'IB n'utilise pas de listes de questions.
Marge de tolérance	Écart avec la note définitive des examinateurs principaux qui, selon l'IB, est assez minime pour montrer que la notation de l'examinateur respecte bien la norme. Les marges de tolérance sont nécessaires, car la notation relève du jugement, et même les correcteurs expérimentés attribueront des notes légèrement différentes en recarrant un même travail. Les marges de tolérance varient en fonction du nombre de points, du type de question et de la matière.
Matériel de soutien pédagogique	Renseignements complémentaires destinés à aider les enseignants à comprendre les exigences du cours de l'IB. Il a pour but de fournir des conseils pratiques et de faciliter ainsi la compréhension et la mise en œuvre de la théorie présentée dans les guides pédagogiques.
Matière pilote	Matière en phase d'essai qui, en cas de succès, sera proposée à tous.
Mauvaise administration	Action d'une école du monde de l'IB qui enfreint les règlements de l'IB et peut menacer l'intégrité des examens et des évaluations de l'organisation. Elle peut avoir lieu avant, pendant ou après l'évaluation ou l'examen.
Mauvaise conduite	La mauvaise conduite est un comportement (qu'il soit délibéré ou fortuit) procurant ou susceptible de procurer un avantage déloyal au candidat qui l'adopte ou à tout autre candidat dans une ou plusieurs composantes d'évaluation. Un comportement susceptible de porter préjudice à un autre candidat est également considéré comme une mauvaise conduite
Mesures compensatoires	Un candidat affecté par des circonstances défavorables peut bénéficier de mesures compensatoires, si cela n'avantage pas le candidat concerné par rapport aux autres. Dans pareil cas, sa note finale pour la ou les disciplines concernées sera augmentée si un ou deux points séparent son total de points du seuil d'attribution de la note finale supérieure. Il s'agit du seul aménagement possible pour les candidats affectés par des circonstances défavorables. Si le total des points du candidat ne suit pas l'écart requis, aucun ajustement ne pourra être apporté.
Modèle d'assurance de la qualité	Approche adoptée par l'IB pour s'assurer que les candidats reçoivent des résultats corrects à l'issue de l'évaluation. L'examinateur principal établit la norme de notation correcte pour chaque question et

Terme	Définition
	chaque examinateur doit reproduire cette norme. Pour les évaluations notées en externe, cela s'effectue en prodiguant des conseils aux examinateurs durant la normalisation, en vérifiant leur compréhension de la norme au moyen de copies de validation, puis en contrôlant régulièrement leur notation au moyen des copies de contrôle.
Mot-consigne	Le ou les mots dans une question qui expliquent l'objectif d'évaluation qui est évalué.
Niveau	Niveau donné lorsque le travail du candidat reflète le descripteur correspondant. Les niveaux sont indiqués dans la colonne de gauche des critères d'évaluation.
Nombre de points	Valeur reflétant la qualité de la réponse du candidat à la question posée.
Normalisation	Processus collaboratif par lequel une norme d'évaluation commune est établie entre les réviseurs de notation ou les examinateurs.
Normalisation interne	Processus par lequel tous les enseignants d'une matière d'un même établissement s'assurent d'appliquer la même norme de notation.
Norme	Accomplissement attendu pour obtenir une note, une note finale ou un résultat particulier.
Notation électronique	Processus par lequel les examinateurs notent le matériel d'examen directement sur leur écran d'ordinateur.
Note définitive	Note attribuée au travail d'un candidat par l'examineur principal. Il s'agit de la note que tout autre examinateur doit chercher à reproduire en corrigeant ce travail. (Voir Modèle d'assurance de la qualité .)
Note finale	Description de l'accomplissement du candidat. Les notes finales pour les travaux des candidats vont de 1 (la plus faible) à 7 (la plus élevée). La note finale représente le jugement émis par l'IB sur les qualités générales mises en évidence par le candidat. Elle est cohérente d'une année à l'autre et d'une matière à l'autre.
Objectif spécifique	Un des énoncés décrivant les compétences, les connaissances et les compréhensions qui seront évaluées.
Outil d'évaluation	Méthode de recueil d'informations sur l'accomplissement et la compréhension de l'apprenant.
Outil de familiarisation	Simulation générique d'un examen que les candidats peuvent passer afin d'apprendre à utiliser

Terme	Définition
	l'environnement et les outils d'un examen sur ordinateur.
Pertinence de construit	Mesure dans laquelle l'évaluation permet effectivement de contrôler les compétences et les connaissances qu'elle est supposée évaluer. Un exemple de faible pertinence de construit serait de contrôler les compétences pratiques d'un élève au moyen d'un examen écrit.
Plagiat	Présentation, intentionnelle ou non, des idées, des propos ou du travail d'une autre personne sans mentionner correctement, clairement et explicitement les sources correspondantes.
Portfolios électroniques	Système ou processus permettant aux établissements de charger les examens ou les travaux de cours évalués en interne afin que la notation soit révisée en externe par l'IB.
Prévisibilité	Le terme « prévisibilité » désigne la capacité à évaluer ce qui se passera ou à quel moment une chose se passera. Dans le contexte de l'évaluation, cela fait référence à l'aptitude des établissements scolaires à anticiper les questions qui seront posées dans une épreuve et le moment auquel elles seront posées. Il est fondamental que les pratiques de travail de l'IB en matière d'évaluation comportent une bonne prévisibilité, car, en adhérant à ce principe, l'IB indique qu'il respecte les exigences de ses construits, qui sont publiées à l'intention des enseignants, ce qui conduit à une possibilité d'évaluation « juste » du point de vue de l'alignement avec le programme d'études (par exemple, ce que l'IB a annoncé vouloir évaluer est évalué).
Procédure pour note finale manquante	Mécanisme visant à attribuer une note finale à un candidat lorsque l'IB ne peut obtenir une note finale exacte ou juste à partir du travail effectué par ce candidat. Elle est pertinente dans les cas où le manque de preuves est dû aux actions de l'IB ou de tiers (à l'exclusion de l'établissement) et où il ne serait pas raisonnable de demander au candidat de recommencer l'évaluation.
Procédure pour note manquante	Mécanisme visant à attribuer une note à un candidat lorsque l'IB ne peut obtenir une note exacte ou juste à partir du travail effectué par ce candidat. Elle est pertinente dans les cas où le manque de preuves est dû aux actions de l'IB ou de tiers (à l'exclusion de l'établissement) et où il ne serait pas raisonnable de demander au candidat de recommencer l'évaluation.

Terme	Définition
Publication des résultats	Processus par lequel les candidats reçoivent les notes finales attribuées par l'IB en fonction de leurs évaluations.
QCM	Voir Question à choix multiple .
Question	Tâche ou activité utilisée pour permettre à un candidat de montrer sa compétence dans une matière.
Question à choix multiple	Question dans laquelle un candidat doit choisir la bonne réponse dans une liste de réponses possibles fournie avec la question.
Réclamation concernant les résultats	Réexamen de la notation (niveaux/points) entreprise à la demande d'un établissement.
Recorrection des réponses électroniques attribuées à un examinateur	Processus consistant à recorriger les réponses électroniques attribuées à un examinateur lorsque sa notation est jugée incohérente ou qu'elle s'écarte considérablement de la norme requise. Cela se produit généralement lorsqu'un examinateur ne passe pas avec succès la révision de notation.
Remise du travail	Pour le candidat (ou l'établissement au nom du candidat), action de rendre la version définitive d'un travail à l'IB.
Réponse à l'évaluation	Terme utilisé pour décrire tout matériel produit par un candidat en réponse au matériel d'évaluation.
Réponse atypique	Une réponse à une tâche qui est très différente des réponses généralement reçues. Il peut s'agir, par exemple, d'un travail incomplet, d'un travail non conforme, d'un travail contenant des réponses inattendues, d'un travail problématique ou d'une fraude.
Reprise	Seconde tentative ou tentative ultérieure consistant à se représenter à un ou plusieurs examens dans l'espoir d'obtenir le certificat du PEI ou le diplôme de l'IB, ou d'augmenter le total des points sur un certificat du PEI ou un diplôme de l'IB déjà reçu.
Résultats de cours	Les résultats de cours constituent le principal document présentant les résultats. Ce document montre chaque discipline étudiée par le candidat ainsi que la note finale obtenue (1-7). Il présente également la note finale obtenue pour le tronc commun composantes et l'évaluation interdisciplinaire, et atteste que le candidat a satisfait aux exigences de l'établissement scolaire concernant le service communautaire.
Réunion de normalisation	Réunion tenue par les examinateurs principaux, qui vise à décrire la norme de notation requise et à produire des copies de contrôle.

Terme	Définition
Révision de notation	Processus destiné à s'assurer qu'une norme de notation commune est utilisée, en révisant la notation d'échantillons de travaux de candidats évalués et en ajustant, le cas échéant, les évaluations du correcteur.
Révision de notation externe	Voir Révision de notation .
Session	Voir Session d'examens .
Session d'examens	Période pendant laquelle les examens sont passés et notés. Chaque année, l'IB organise deux sessions d'examens : une en mai et une en novembre.
Seuil d'attribution des notes finales	Point où l'accomplissement du candidat passe d'une note finale à une autre. Il est souvent utilisé pour indiquer le total des niveaux par critère le plus bas ou le plus haut pour l'attribution d'une note finale donnée.
Stratégie d'évaluation	Méthode ou approche adoptée pour recueillir des informations relatives à l'apprentissage des élèves (par exemple, des observations, des tâches ouvertes ou des réponses choisies).
Surveillant d'examen	Fonction de la personne qui surveille et contrôle l'environnement dans lequel se déroule l'examen.
Tâche d'évaluation	Activité ou série d'activités auxquelles participent les élèves en vue de l'évaluation.
Totalisation	Processus consistant à additionner les points et les notes pour obtenir un résultat final.
Validité	Terme générique décrivant la pertinence d'une évaluation ou du but pour lequel les résultats de l'évaluation sont utilisés.