

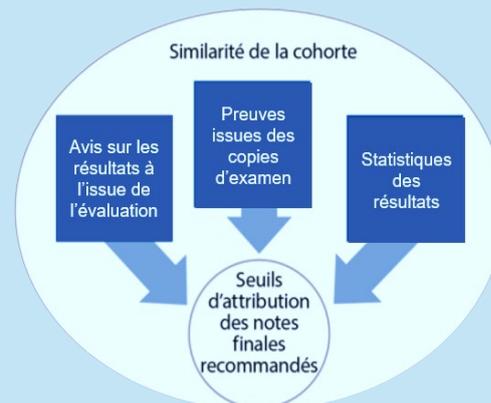
Recours aux statistiques lors de la notation : contextualisation ou conditionnement ?

[Article intégral \(en anglais\) – « Statistically guided grading judgements: contextualisation or contamination? » \(tandfonline.com\)¹](#)

INTRODUCTION

Plusieurs sources de preuves tirées de l'évaluation sont examinées au cours du processus d'attribution des notes finales de l'IB, qui consiste à convertir les notes en notes finales et à garantir des résultats justes aux élèves. Les preuves qualitatives et quantitatives sont analysées afin de déterminer les seuils d'attribution des notes finales, et les preuves statistiques sont confrontées au jugement de l'examinateur ou de l'examinatrice, ainsi qu'au retour d'information de l'équipe enseignante sur les examens. Au printemps 2022, l'IB a effectué une mise à l'essai pour déterminer dans quelle mesure les décisions de notation des examinateurs et examinatrices étaient influencées par la consultation de preuves statistiques et du retour d'information de l'équipe enseignante. L'objectif était de comparer différentes approches en matière d'attribution des notes finales de l'IB et de déterminer si les résultats et les expériences de notation variaient de manière substantielle en l'absence de preuves tirées de l'évaluation supplémentaires avant la notation des copies.

Preuves venant étayer le choix des seuils d'attribution des notes finales



(IB, 2019²)

CONTEXTE

Dans le cadre de l'attribution des notes finales, l'IB s'appuie sur une série de preuves pour déterminer les seuils d'attribution des notes finales les plus adaptés pour chaque matière et pour chaque examen.

Par le passé, les examinateurs et examinatrices recevaient des preuves tirées de l'évaluation avant la notation des copies dans la plupart des matières (par exemple, les seuils recommandés statistiquement, les notes moyennes, la distribution des notes, etc.). Ce procédé d'attribution des notes finales correspondait au « modèle enrichi ».

Ce modèle a cependant été remplacé en mai 2019 par un « modèle limité » dans certaines matières : les examinateurs et examinatrices devaient alors noter les copies d'examen sans pouvoir consulter ces preuves supplémentaires.

Modèle enrichi

Les examinateurs et examinatrices consultent les données statistiques et le retour d'information de l'équipe enseignante, puis notent les copies d'examen.



Modèle limité

Les examinateurs et examinatrices notent les copies d'examen sans consulter les données statistiques ni le retour d'information de l'équipe enseignante.



OBJECTIFS

Cette étude avait pour objectif principal de comparer les résultats de notation et les expériences des examinateurs et examinatrices dans le cadre du modèle enrichi et du modèle limité d'attribution des notes finales. Il s'agissait notamment de mesurer l'intégration des différentes sources de preuves dans l'attribution des notes finales de l'IB et l'influence de la consultation préalable de preuves contextuelles sur la décision de notation des examinateurs et examinatrices.

Cette étude a été guidée par les questions de recherche suivantes.

1. Comment et dans quel but les preuves reposant sur le jugement et les preuves statistiques sont-elles combinées au cours de l'attribution des notes finales de l'IB ?
2. Dans quelle mesure les preuves statistiques relatives aux examens ont-elles une incidence sur les décisions de notation des examinateurs et examinatrices ?
3. Les attributions de notes finales avec et sans consultation préalable de preuves statistiques conduisent-elles à des résultats de notation similaires ?
4. Quels sont les avantages et les inconvénients perçus par les examinateurs et examinatrices qui consultent des preuves statistiques sur les examens dans le cadre de l'attribution des notes finales de l'IB ?

MÉTHODOLOGIE

Les processus d'attribution des notes finales ont été recréés dans le cadre de neuf examens concernant cinq matières. Au total, 30 personnes ont participé à l'étude (25 examinateurs et examinatrices, 5 responsables de matière) et tout le matériel d'évaluation était issu de la session de mai 2019.

Pour chaque matière, le modèle d'attribution des notes finales a été inversé par rapport à celui utilisé en mai 2019.

Matière	Modèle enrichi	Modèle limité
Anglais A : littérature (NS) – Épreuve 1 et épreuve 2	Essai	Mai 2019
Japonais A : littérature (NS) – Épreuve 1 et épreuve 2		
Espagnol A : littérature (NS) – Épreuve 1 et épreuve 2		
Gestion des entreprises NS – Épreuve 1 et épreuve 2	Mai 2019	Essai
Évaluation électronique de mathématiques du PEI		

L'étude repose sur l'utilisation de méthodes mixtes, notamment la réalisation de comparaisons quantitatives des résultats de notation et l'organisation de groupes de réflexion pour collecter les commentaires des personnes participantes sur leur expérience.

LIMITES

- Petit nombre de matières : il est possible de combiner et d'évaluer les preuves différemment dans d'autres contextes disciplinaires.
- Risque potentiel que certaines personnes participantes aient retenu les seuils d'attribution des notes finales initiaux.

Compte tenu de ces faiblesses, cette étude avait pour objectif principal d'encourager la discussion sur la meilleure façon de combiner les différentes formes de preuves dans le cadre de l'attribution des notes finales, plutôt que de fournir des preuves définitives en faveur d'une approche en particulier.

RÉSULTATS

Quantitatifs

Les conclusions préliminaires suggèrent que les deux approches conduisent à des résultats de notation généralement comparables. Les copies jugées dignes de recevoir une note finale étaient alignées en proportion similaire sur les seuils recommandés statistiquement dans les deux modèles. Seule exception : la littérature japonaise, pour laquelle la divergence s'explique probablement par la très faible taille de la cohorte et la quantité limitée de copies d'examen permettant la prise de décision.

Qualitatifs

Les groupes de réflexion ont fait ressortir cinq thèmes principaux.

1. Les preuves issues des copies d'examen sont considérées comme essentielles pour le rôle de l'examineur ou de l'examinatrice lors de l'attribution des notes finales.
2. Les recommandations statistiques peuvent obscurcir le jugement de l'examineur ou de l'examinatrice au cours de la notation.
3. Les preuves statistiques les plus pertinentes pour guider l'examineur ou l'examinatrice dans la notation sont les données relatives aux éléments de questions.
4. Les preuves statistiques sont utiles à maints égards pour les examinateurs et examinatrices, outre la notation.
5. Il n'existe pas de formule universelle : les besoins et les exigences varient d'une matière à l'autre.

RECOMMANDATIONS

- 1. Données relatives aux éléments de questions :** continuer à fournir des données relatives aux éléments de questions lorsqu'elles sont disponibles, pour aider les examinateurs, les examinatrices et les responsables de matière à reconnaître les principales questions discriminantes pour la détermination des seuils d'attribution des notes finales. Étudier la possibilité de collecter des données relatives aux éléments de questions dans les matières ou les composantes où elles ne sont pas disponibles actuellement (par exemple, pour l'épreuve 2 de langue A).
- 2. Calendrier d'attribution :** envisager de partager d'autres données statistiques avec l'examineur superviseur ou l'examinatrice superviseuse après la session, plutôt qu'avant la notation, afin de limiter le risque d'obscurcir le jugement lors de la notation et de réduire la charge administrative des responsables de matière.
- 3. Flexibilité par matière :** chercher à trouver l'équilibre entre la nécessité d'instaurer des processus systématiques d'une part, et les exigences et demandes relatives aux différentes matières d'autre part (par exemple, en raison des différences de tailles de cohortes, des gammes de points ou de la langue d'enseignement).
- 4. Communication avec les examinateurs superviseurs et les examinatrices superviseuses :** afin que les examinateurs et les examinatrices se sentent partie intégrante du processus, veiller à communiquer les décisions relatives aux seuils d'attribution des notes finales, à souligner l'importance de leur contribution et à expliquer de quelle façon les différentes formes de preuves ont été combinées pour garantir les résultats les plus justes pour les élèves.
- 5. Pertinence pour les autres tâches des examinateurs et examinatrices :** comme les preuves tirées de l'évaluation sont utiles pour d'autres activités, telles que l'élaboration des épreuves, la rédaction des rapports pédagogiques et la préservation des normes de notation, il pourrait également être utile de les partager avec d'autres examinateurs et examinatrices en dehors de l'équipe chargée de l'attribution des notes finales.
- 6. Poursuite de la recherche :** pour approfondir cette recherche, il est possible de reproduire l'étude dans d'autres disciplines ou d'explorer l'incidence de la consultation des recommandations statistiques par les examinateurs et les examinatrices après plutôt qu'avant la notation. Il pourrait par ailleurs s'avérer particulièrement utile d'analyser les répercussions sur les matières réunissant des cohortes très restreintes, dans la mesure où les preuves statistiques sont, par nature, moins fiables dans ces contextes.

¹ BADHAM, L. Statistically guided grading judgements: contextualisation or contamination? *Oxford Review of Education*, 2023. DOI : 10.1080/03054985.2023.2290640.

² ORGANISATION DU BACCALAURÉAT INTERNATIONAL [IB]. 2019. *Principes et pratiques de l'évaluation – Des évaluations de qualité à l'ère du numérique*. Cardiff, Royaume-Uni : Organisation du Baccalauréat International.