

El uso de datos estadísticos en la concesión de calificaciones finales: ¿ayudan a contextualizar o entorpecen?

Artículo completo en inglés: [Statistically guided grading judgements: contextualisation or contamination?](https://tandfonline.com) (tandfonline.com)¹

Introducción

Durante la concesión de calificaciones finales del Bachillerato Internacional (IB) se revisan diferentes fuentes de pruebas de evaluación a fin de convertir las puntuaciones en calificaciones y garantizar que el alumnado obtenga resultados justos. Las pruebas cualitativas y cuantitativas se analizan para determinar los límites de calificación, y las pruebas estadísticas se sopesan con las valoraciones del examinador/a y los comentarios del personal docente sobre los exámenes. En el segundo trimestre de 2022, el IB realizó un estudio para explorar cómo el acceso a pruebas estadísticas y a comentarios del personal docente influía en las decisiones del equipo examinador sobre la concesión de calificaciones. El objetivo era comparar diferentes enfoques de concesión de calificaciones finales y determinar si los resultados y las experiencias variaban de manera considerable si el equipo examinador no tenía acceso a pruebas de evaluación adicionales antes de corregir los exámenes.



Contexto

En la concesión de calificaciones finales, el IB utiliza distintos tipos de pruebas a fin de determinar los límites de calificación más adecuados para cada asignatura y examen.

Tradicionalmente, en la mayoría de las asignaturas, el equipo examinador utilizaba un **modelo ampliado** para la concesión de calificaciones, en el que revisaba pruebas de evaluación antes de corregir los exámenes (p. ej., límites recomendados estadísticamente, puntuaciones medias, distribuciones de las puntuaciones, etc.).

Sin embargo, en mayo de 2019, se introdujo un **modelo reducido** en algunas asignaturas, en el que el equipo examinador corrige los exámenes sin tener acceso a pruebas adicionales.

Modelo ampliado

El equipo examinador revisa datos estadísticos y comentarios del personal docente antes de corregir los exámenes.



Modelo reducido

El equipo examinador corrige los exámenes sin revisar datos estadísticos ni comentarios del personal docente.



Objetivos

El objetivo general de este estudio era comparar los resultados de la concesión de calificaciones finales y las experiencias del equipo examinador con el modelo ampliado y el modelo reducido. Se investigó cómo las diferentes fuentes de pruebas se integran en la concesión de calificaciones finales del IB y cómo la revisión de pruebas contextuales antes de la valoración de las calificaciones influye en las decisiones del equipo examinador.

En el estudio se exploraron las siguientes preguntas de investigación:

1. ¿Cómo y con qué efecto se combinan las pruebas estadísticas y de valoraciones durante la concesión de calificaciones finales del IB?
2. ¿En qué medida influye el acceso a pruebas estadísticas en las decisiones sobre las calificaciones finales del equipo examinador?
3. Cuando el equipo examinador revisa pruebas estadísticas, ¿se conceden calificaciones similares a cuando no lo hace?
4. ¿Cuáles son las ventajas y desventajas perceptibles de que el equipo examinador revise pruebas estadísticas de los exámenes durante la concesión de calificaciones finales?

Metodología

Los procesos de concesión de calificaciones finales se replicaron en nueve exámenes de cinco asignaturas. En el estudio participaron 30 personas (25 examinadores/as y 5 responsables de las asignaturas), y todos los materiales de evaluación se tomaron de la convocatoria de mayo de 2019.

Cada asignatura siguió un modelo de concesión de calificaciones contrario al de mayo de 2019.

Asignatura	Modelo ampliado	Modelo reducido
Prueba 1 y 2 de Inglés A: Literatura NS	Estudio	Convocatoria de mayo de 2019
Prueba 1 y 2 de Japonés A: Literatura NS		
Prueba 1 y 2 de Español A: Literatura NS		
Prueba 1 y 2 de Gestión Empresarial NS	Convocatoria de mayo de 2019	Estudio
Evaluación electrónica de Matemáticas del PAI		

Se empleó un enfoque de métodos mixtos, donde se realizaron comparaciones cuantitativas de las calificaciones finales concedidas y se llevaron a cabo grupos de discusión para recabar comentarios sobre las experiencias de quienes participaron.

Limitaciones

- Número reducido de asignaturas. Las pruebas pueden combinarse y valorarse de forma diferente en otros contextos disciplinarios.
- Posible riesgo de que algunos/as participantes hayan recordado los límites de calificación originales.

Debido a estas limitaciones, el objetivo principal de este estudio era fomentar el debate sobre la mejor manera de combinar diferentes tipos de pruebas en la concesión de calificaciones finales, en lugar de aportar pruebas concluyentes a favor de un enfoque concreto.

Hallazgos

Cuantitativos

Los hallazgos preliminares indican que ambos enfoques conducen, en general, a resultados de calificación comparables. La proporción de exámenes que se consideraba que debía calificarse se ajustaba de forma similar a los límites recomendados estadísticamente en ambos modelos. La única excepción fue Japonés A: Literatura, donde la discrepancia probablemente se debe a que el grupo de estudiantes era muy pequeño y había una disponibilidad limitada de exámenes para la toma de decisiones.

Cualitativos

Surgieron cinco temas principales de los grupos de discusión:

1. Las pruebas que aportan los exámenes son fundamentales para el equipo examinador durante la concesión de calificaciones finales.
2. Las recomendaciones estadísticas pueden entorpecer la valoración del examinador/a en la concesión de calificaciones finales.
3. Al conceder calificaciones, los datos estadísticos que resultan más útiles para el equipo examinador son los de los datos pertinentes de elementos de preguntas.
4. El equipo examinador puede usar las pruebas estadísticas de muchas otras maneras fuera del ámbito de la concesión de calificaciones.
5. Una misma solución no se ajusta a todos los casos. Las necesidades y requisitos varían según la asignatura.

Recomendaciones

- 1. Datos de elementos de preguntas:** Seguir proporcionando datos de elementos de preguntas al equipo examinador en las asignaturas en las que sea posible, a fin de ayudar a examinadores/as y responsables de las asignaturas a identificar las preguntas que sirven para la discriminación en el establecimiento de límites de calificación. Investigar la posibilidad de obtener datos pertinentes de elementos de preguntas en asignaturas y componentes en los que actualmente no se dispone de ellos (p. ej., prueba 2 de Lengua A).
- 2. Cronograma de concesión de calificaciones:** Considerar la posibilidad de compartir otros datos estadísticos con el personal examinador supervisor después de la convocatoria en lugar de antes de corregir los exámenes, a fin de minimizar el riesgo de entorpecer las valoraciones de los examinadores/as durante la concesión de calificaciones y reducir la carga de trabajo administrativo de las personas responsables de las asignaturas.
- 3. Flexibilidad por asignatura:** Intentar equilibrar la necesidad de tener procesos coherentes con los requisitos y exigencias de cada asignatura (p. ej., aquellos que surgen de diferencias en el tamaño de los grupos de estudiantes, en los rangos de puntuaciones o en las lenguas de instrucción).
- 4. Comunicación con el equipo examinador supervisor:** Intentar comunicarle al equipo examinador supervisor —a fin de que no sienta que se le excluye— las decisiones finales sobre el establecimiento de los límites de calificación, así como el modo en que sus aportes contribuyeron a estas decisiones y la manera en que se han combinado los diferentes tipos de pruebas para garantizar que el alumnado reciba resultados justos.
- 5. Pertinencia para otras tareas del equipo examinador:** Las pruebas de las evaluaciones son útiles para otras tareas de los examinadores/as (como la preparación de los exámenes, la redacción de los informes generales de las asignaturas y el mantenimiento de los estándares de corrección). Asimismo, puede ser útil compartirlas con examinadores/as supervisores con más experiencia que no forman parte del equipo de concesión de calificaciones finales.
- 6. Futuras investigaciones:** Entre las posibilidades para futuras investigaciones se incluyen replicar el estudio en diferentes áreas disciplinarias o analizar el efecto de que el equipo examinador revise las recomendaciones estadísticas después de conceder las calificaciones en lugar de antes. Las investigaciones sobre el impacto en grupos de estudiantes muy pequeños también pueden ser útiles, ya que las pruebas estadísticas son intrínsecamente menos fiables en estos contextos.

¹ BADHAM, L. "Statistically guided grading judgements: contextualisation or contamination?". En *Oxford Review of Education* [en línea]. 2023. <10.1080/03054985.2023.2290640>.

² BACHILLERATO INTERNACIONAL. *Principios y prácticas de evaluación del IB: evaluaciones de calidad en la era digital*. Cardiff (Reino Unido), 2019.