



International Baccalaureate®
Baccalauréat International
Bachillerato Internacional

Principios y prácticas de evaluación del IB: una evaluación de calidad en la era digital



International Baccalaureate®
Baccalauréat International
Bachillerato Internacional

Principios y prácticas de evaluación del IB: una evaluación de calidad en la era digital

Principios y prácticas de evaluación del IB: una evaluación de calidad en la era digital

Versión en español del documento publicado en mayo de 2026 con el título
Assessment principles and practices: Quality assessment in a digital age

Publicada en mayo de 2026

Publicada por la Organización del Bachillerato Internacional, una fundación educativa sin fines de lucro con sede en Rue du Pré-de-la-Bichette 1, 1202 Ginebra (Suiza)

Sitio web: ibo.org/es

© Organización del Bachillerato Internacional, 2026

La Organización del Bachillerato Internacional (conocida como IB) ofrece cuatro programas educativos exigentes y de calidad a una comunidad de colegios de todo el mundo, con el propósito de crear un mundo mejor y más pacífico. Esta publicación forma parte de una gama de materiales producidos con el fin de apoyar dichos programas.

El IB puede utilizar diversas fuentes en su trabajo y comprueba la información para verificar su exactitud y autoría original, en especial al hacer uso de fuentes de conocimiento comunitario, como Wikipedia. El IB respeta la propiedad intelectual, y hace denodados esfuerzos por identificar a las personas titulares de los derechos y obtener la debida autorización antes de la publicación de todo material protegido por derechos de autor utilizado. El IB agradece las autorizaciones recibidas para utilizar los materiales incluidos en esta publicación y enmendará cualquier error u omisión lo antes posible.

El IB pretende que el español utilizado en sus publicaciones sea comprensible para la totalidad de hablantes de esta lengua y no refleje una variante particular o regional. Asimismo, en la redacción se aplica un enfoque de comunicación no sexista, que constituye una etapa intermedia dentro de la actual trayectoria global del IB respecto al acceso y la inclusión.

Todos los derechos reservados. Esta publicación no puede reproducirse, almacenarse en un sistema de archivo y recuperación de datos ni distribuirse de forma total o parcial, de manera alguna ni por ningún medio, sin la previa autorización por escrito del IB o sin que esté expresamente permitido en la [normativa de uso de la propiedad intelectual del IB](#).

Los artículos promocionales y las publicaciones del IB pueden adquirirse en la [tienda virtual del IB](#) (correo electrónico: sales@ibo.org). Está prohibido el uso comercial de las publicaciones del IB (tanto las incluidas en las tasas como las que se pueden adquirir por separado) por parte de terceras personas que actúen en el entorno de la Organización del Bachillerato Internacional sin haber establecido una relación formal con ella (incluidos, entre otros, organizaciones que imparten clases, empresas proveedoras de desarrollo profesional, empresas editoriales del sector educativo y compañías que ofrecen servicios de planificación curricular o plataformas digitales que brindan recursos pedagógicos). Dicho uso comercial solo está permitido con la correspondiente licencia por escrito otorgada por el IB. Las solicitudes de licencias deben enviarse a copyright@ibo.org. Encontrará más información al respecto en el [sitio web del IB](#).

Declaración de principios del IB

El Bachillerato Internacional tiene como meta formar personas solidarias, informadas y ávidas de conocimiento, capaces de contribuir a crear un mundo mejor y más pacífico, en el marco del entendimiento mutuo y el respeto intercultural.

En pos de este objetivo, la organización colabora con establecimientos escolares, gobiernos y organizaciones internacionales para crear y desarrollar programas de educación internacional exigentes y métodos de evaluación rigurosos.

Estos programas alientan a estudiantes del mundo entero a adoptar una actitud activa de aprendizaje durante toda su vida, a actuar de forma compasiva y a entender que otras personas, con sus diferencias, también pueden estar en lo cierto.



Perfil de la comunidad de aprendizaje del IB

El objetivo fundamental de los programas del Bachillerato Internacional (IB) es formar personas con mentalidad internacional que, conscientes de la condición que las une como seres humanos y de la responsabilidad que comparten de velar por el planeta, contribuyan a crear un mundo mejor y más pacífico.

Como miembros de la comunidad de aprendizaje del IB, nos esforzamos por demostrar los siguientes atributos:

INDAGACIÓN

Cultivamos nuestra curiosidad, a la vez que desarrollamos habilidades para la indagación y la investigación. Sabemos cómo aprender de manera autónoma y junto con otras personas. Aprendemos con entusiasmo y mantenemos estas ansias de aprender durante toda la vida.

CONOCIMIENTO

Desarrollamos y usamos nuestra comprensión conceptual mediante la exploración del conocimiento en una variedad de disciplinas. Nos comprometemos con ideas y cuestiones de importancia local y mundial.

RAZONAMIENTO

Utilizamos habilidades de pensamiento crítico y creativo para analizar y proceder de manera responsable ante problemas complejos. Actuamos por propia iniciativa al tomar decisiones razonadas y éticas.

COMUNICACIÓN

Nos expresamos con confianza y creatividad en diversas lenguas, lenguajes y maneras. Colaboramos eficazmente, escuchando atentamente las perspectivas de otras personas y grupos.

INTEGRIDAD

Actuamos con integridad y honradez, con un profundo sentido de la equidad, la justicia y el respeto por la dignidad y los derechos de las personas en todo el mundo. Asumimos la responsabilidad de nuestros propios actos y sus consecuencias.

MENTALIDAD ABIERTA

Desarrollamos una apreciación crítica de nuestras propias culturas e historias personales, así como de los valores y tradiciones de otras personas. Buscamos y consideramos distintos puntos de vista y mostramos disposición a aprender de la experiencia.

SOLIDARIDAD

Mostramos empatía, sensibilidad y respeto. Nos comprometemos a ayudar y actuamos con el propósito de influir positivamente en la vida de las personas y el mundo que nos rodea.

AUDACIA

Abordamos la incertidumbre con previsión y determinación. Trabajamos de manera autónoma y colaborativa para explorar nuevas ideas y estrategias innovadoras. Mostramos ingenio y resiliencia cuando enfrentamos cambios y desafíos.

EQUILIBRIO

Entendemos la importancia del equilibrio físico, mental y emocional para lograr el bienestar propio y el de las demás personas. Reconocemos nuestra interdependencia con respecto a otras personas y al mundo en que vivimos.

REFLEXIÓN

Evaluamos detenidamente el mundo y nuestras propias ideas y experiencias. Nos esforzamos por comprender nuestras fortalezas y debilidades para, de este modo, contribuir a nuestro aprendizaje y desarrollo personal.

El perfil de la comunidad de aprendizaje engloba diez atributos valorados por los Colegios del Mundo del IB. Tenemos la convicción de que estos atributos, y otros similares, pueden ayudar a personas y grupos a ser miembros responsables de las comunidades locales, nacionales y mundiales.

Índice

Introducción y descripción general	1
Introducción	1
Recursos adicionales	3
Uso de este recurso	5
Principios y prácticas de evaluación	11
Buena evaluación	12
El papel de la tecnología en la evaluación del IB	21
Sección A: Principios de evaluación	25
La evaluación en la educación	25
Validez	30
Elementos de la cadena de validez	33
Estándares	50
Descripción del éxito del alumnado en la evaluación sumativa	54
Corrección de las evaluaciones del IB	57
Fomento de una mentalidad ética	60
Sección B: Las prácticas de evaluación del IB	63
Definición de las prácticas	63
Comunicación de los logros del alumnado	64
El proceso de evaluación: funciones y responsabilidades	72
Integridad de la evaluación	76
Priorización de la equidad para todo el alumnado	80
El ciclo de vida de la evaluación	84
Elaboración de evaluaciones	86
Corrección	93
Moderación	103
Moderación y expectativas del profesorado	105
Concesión de calificaciones finales y totalización	110
Controles de calidad	119
El Comité de la Evaluación Final	120
Preparación para la publicación de los resultados	122
Consultas sobre los resultados y apelaciones sobre la evaluación	125
Comentarios para los colegios	131
Sección C: Procesos específicos de los programas del IB	133
Definición de los procesos específicos de los programas	133
Elementos comunes a todos los programas	134

El perfil de la comunidad de aprendizaje del IB	136
Necesidades y soluciones específicas de los programas	137
Programa del Diploma	138
Programa de Orientación Profesional	143
Programa de los Años Intermedios	147
Programa de la Escuela Primaria	151

Apéndices **154**

La moderación de la evaluación interna en mayor detalle	154
Bibliografía	160
Glosario	163
Recursos para imprimir	175

Introducción

Los programas del Bachillerato Internacional (IB) se imparten en más de 150 países y regiones o territorios que abarcan una amplia variedad de tradiciones y contextos educativos. A algunas personas, la filosofía y los métodos de evaluación del IB les resultarán familiares; a otras, el sistema les puede parecer complejo y desconocido.

El IB asume todo un desafío mediante sus principios de evaluación, ya que se compromete a evaluar lo que es verdaderamente importante, en lugar de lo que es simplemente fácil de medir. Alec Peterson, el primer director general del IB, expresó este desafío dentro del ecosistema educativo como un cuestión de equilibrio:

Lo que se necesita es un proceso de evaluación que sea lo más válido posible, en el sentido de que realmente evalúe las capacidades y la personalidad del alumno con relación a la próxima etapa de su vida, pero [que sea] a la vez lo suficientemente fiable como para garantizar a alumnos, padres, profesores e instituciones receptoras que se está haciendo justicia. Sin embargo, dicho proceso no debe, por su "efecto de repercusión", distorsionar la buena enseñanza ni ser demasiado lento, ni absorber demasiados de nuestros escasos recursos educativos.






(Peterson, 1971, pp. 27-55)

El objetivo de esta publicación es dilucidar los principios que el IB emplea para asegurarse de que la evaluación en sus programas sea significativa, justa y beneficiosa para el alumnado, de una manera auténtica y pertinente que refleje el proceso de aprendizaje y enseñanza.

A grandes rasgos, la evaluación en el IB se divide en dos tipos, como se muestra en la figura 1.

Figura 1

Evaluación sumativa y evaluación formativa

El propósito		Evaluación sumativa: demostrar dominio o medir la capacidad del alumnado Evaluación formativa: identificar áreas de desarrollo o áreas en las que centrarse (como clase o individualmente)
El resultado		Evaluación sumativa: un producto Evaluación formativa: un proceso
El momento de la evaluación		Evaluación sumativa: culminación del aprendizaje Evaluación formativa: continua a lo largo del aprendizaje
El modo de aportar pruebas		Evaluación sumativa: carpetas, exámenes, y trabajos de clase (normalmente individuales) Evaluación formativa: observaciones, tarjetas de salida, compartir en parejas, y autoevaluación (normalmente colaborativa)
La herramienta de evaluación		Evaluación sumativa: criterios y/o bandas de puntuación Evaluación formativa: tablas de evaluación que sirven de andamiaje para el desarrollo y la progresión del aprendizaje

Comprender las distintas funciones de estas evaluaciones es fundamental para la comunidad del IB.

Los resultados de la evaluación tienen un impacto significativo en la vida del alumnado, ya que influyen en su progresión hacia estudios posteriores o carreras profesionales. La evaluación es una herramienta que

puede dar lugar a resultados tanto positivos como negativos. Como docentes, padres, madres y tutores/as legales, o estudiantes, es fundamental comprender los puntos fuertes, los puntos débiles y las decisiones que hay detrás de las evaluaciones. En esta publicación se abordan preguntas e inquietudes habituales, tales como:

- “¿Y por qué hace eso el IB?”: comprender los principios en los que se basan las prácticas de evaluación del IB
- “Pero esto no es justo, necesitaba una calificación más alta”: explicar la equidad y las implicaciones de las decisiones de evaluación
- “Nunca antes me lo había planteado así”: desglosar el proceso de evaluación
- “Solamente quiero utilizar las evaluaciones para poder enseñar mejor”: destacar la importancia de las evaluaciones de alta calidad con fines tanto formativos como sumativos

Los cambios tecnológicos son constantes, y adoptar la tecnología en las evaluaciones del IB es fundamental para que estas sigan siendo auténticas y reflejen las formas en que el alumnado trabaja, recibe enseñanza y aprende. El hecho incluir el lema “una evaluación de calidad en la era digital” en el título de este recurso sirve para admitir que, si queremos que las evaluaciones del IB continúen siendo pertinentes, deben realizarse en condiciones que el alumnado conozca bien y en las que se sienta cómodo. Aún más importante: así se reconoce que, al hacer la transición del papel a la pantalla, el IB debe permanecer fiel a sus principios fundamentales. Esto implica que la tecnología se utiliza para apoyar el aprendizaje y la evaluación con el fin de crear evaluaciones más válidas. La descripción “una evaluación de calidad en la era digital” subraya el entusiasmo y la capacidad de respuesta del IB ante este proceso y la evolución del panorama, al tiempo que refuerza su compromiso con los principios fundamentales de evaluación del IB.

Esta publicación explica en detalle los principios de evaluación del IB en relación con sus cuatro programas: el Programa de la Escuela Primaria (PEP), el Programa de los Años Intermedios (PAI), el Programa del Diploma (PD), y el Programa de Orientación Profesional (POP). Estos principios abarcan todos los tipos de evaluaciones del IB y explican por qué están diseñadas de la manera en que lo están. Si bien algunas secciones se centran específicamente en la evaluación sumativa, como las evaluaciones entregadas al IB para su corrección externa, los conocimientos expuestos están concebidos para mejorar la práctica educativa en todos los programas. Incluso en contextos donde no se realizan evaluaciones sumativas externas, la comprensión de los elementos que hacen a una evaluación eficaz y justa puede mejorar la experiencia en el aula y preparar mejor al alumnado para futuras evaluaciones. Asimismo, esta publicación busca explicar los principios de evaluación a toda la comunidad del IB, incluido personal docente, personal de dirección de los colegios, personal examinador, y profesionales de admisión universitaria. Está dirigida a las personas que están familiarizadas con el IB y sus evaluaciones, pero que quizás no han profundizado en las teorías subyacentes de la evaluación. Se espera que este recurso ayude a comprender el modo en que los principios de evaluación del IB apoyan su intención educativa más amplia de ofrecer una experiencia de prestigio internacional para todo su alumnado.

Terminología

En las evaluaciones del IB se utilizan abundantes términos con los que tal vez docentes y estudiantes no estén familiarizados. En todas las publicaciones del IB se procura evitar confusiones y mantener la precisión.

Como parte de este compromiso, las abreviaturas y las siglas se escriben completas la primera vez que se utilizan en cada sección de esta publicación. Del mismo modo, los términos importantes se explican la primera vez que se utilizan, ya sea directamente o a través de material al que se puede acceder mediante enlaces.

Además, en esta publicación se ofrece un glosario con una lista completa de términos relacionados con la evaluación. Aparte de los términos generales de evaluación, la terminología específica de cada asignatura se explica en la guía de la asignatura del IB correspondiente.

Recursos adicionales

Este recurso es uno de los muchos publicados por el IB para explicar su enfoque educativo. A veces puede resultar confuso entender cómo estos recursos están relacionados entre sí. En las tablas 1 y 2 se ofrece un resumen de los recursos clave y una explicación de su propósito y función.

Tabla 1

Publicaciones clave del IB para todos los programas

Publicaciones clave del IB: todos los programas	
<i>Principios y prácticas de evaluación del IB: una evaluación de calidad en la era digital</i>	Esta publicación explica el enfoque global que adopta el IB con respecto a la evaluación y cómo pretende aplicarlo en la práctica. Se centra en el elemento de evaluación sumativa (evaluación formal) de la educación del IB.
<i>¿Qué es la educación del IB?</i>	El propósito de esta publicación es comunicar con claridad cuáles son los elementos centrales de la educación del IB. Explica los ideales que constituyen la base de todos los programas del IB. Dado que describe la filosofía educativa del IB, esta publicación también puede servir de apoyo a los colegios a lo largo de su recorrido, que comienza con la adopción del IB y que cubre los muchos años que se espera que sean Colegios del Mundo del IB.
<i>Normas para la implementación de los programas y aplicaciones concretas</i>	Esta publicación contiene un conjunto de criterios que tanto los colegios como el IB han de utilizar para evaluar la eficacia en la implementación de los cuatro programas del IB. Comprende lo siguiente: normas para la implementación de los programas (los requisitos generales que los colegios deben seguir para implementar un programa del IB), aplicaciones concretas (definiciones adicionales de las normas para la implementación que son comunes a todos los programas) y requisitos (específicos de cada programa).

Tabla 2

Publicaciones clave del IB para programas específicos

Publicaciones clave del IB: programas específicos	
<i>El Programa de la Escuela Primaria: de los principios a la práctica</i> <i>El Programa de los Años Intermedios: de los principios a la práctica</i> <i>El Programa del Diploma: de los principios a la práctica</i> <i>El Programa de Orientación Profesional: de los principios a la práctica</i>	Estas publicaciones se centran en el aprendizaje y la enseñanza en el contexto de un programa concreto del IB. Además, explican los requisitos del programa.

Publicaciones clave del IB: programas específicos	
Guías de las asignaturas (solo PAI y PD)	Cada guía contiene información detallada sobre una asignatura, un curso o una disciplina; por ejemplo, Geografía o Artes Visuales. En estas publicaciones se incluyen los objetivos generales y específicos, el programa de estudios y los criterios para la evaluación interna de la asignatura de que se trate. Normalmente incluyen también orientación adicional para el aprendizaje y la enseñanza de cada asignatura.
Continuos de las asignaturas (solo PEP)	Los continuos de las asignaturas son ejemplos que los colegios pueden utilizar para apoyar el diseño de su propia secuenciación de contenidos. Estos pueden adaptarse en función del contexto y los requisitos locales. Los continuos también proporcionan comprensiones conceptuales y ejemplos de conceptos para cada asignatura y sus áreas. Las descripciones generales de las asignaturas que los acompañan muestran claramente los elementos esenciales del PEP en el contexto de las áreas disciplinarias, y aclara el papel de las áreas disciplinarias en un programa transdisciplinario.
<i>Procedimientos de evaluación del Programa de los Años Intermedios</i> <i>Procedimientos de evaluación del Programa del Diploma</i> <i>Procedimientos de evaluación del Programa de Orientación Profesional</i>	Estas publicaciones definen las reglas y procesos específicos que se deben seguir para la evaluación del programa del IB correspondiente. En ellas figuran sus respectivos reglamentos generales, en los que se basan las normas que deben seguir los Colegios del Mundo del IB.
<i>Política de preparación para los exámenes (solo PAI, PD y POP)</i>	En esta publicación se describe la política del IB sobre la preparación para los exámenes, que abarca el almacenamiento seguro de los materiales confidenciales del IB dentro de los colegios y los cambios de fecha u hora de los exámenes.
<i>Instrucciones sobre la realización de los exámenes (solo PD; el POP utiliza la versión del PD) y La realización de los exámenes en pantalla del Programa de los Años Intermedios (PAI)</i>	Esta publicación informa al personal de coordinación y al personal de supervisión de exámenes acerca de las normas relativas a la administración y la realización de los exámenes del programa. Todas las salas de examen deben contar con un ejemplar de esta publicación.

Además de estas valiosas publicaciones, el IB ofrece una variedad de recursos distintos que sirven de orientación para docentes, personal de coordinación y otras partes interesadas. Estos recursos se encuentran en el [sitio web del IB](#) y en el [Centro de Recursos para los Programas](#).

Uso de este recurso

La variedad de temas que se tratan en *Principios y prácticas de evaluación del IB: una evaluación de calidad en la era digital* es muy amplia. Esta publicación de consulta se ha diseñado para utilizarse en función de los intereses y las necesidades de quienes la lean. Las **preguntas sobre el tema** que aparecen en esta sección pueden servir como punto de partida para aprender sobre los aspectos que el lector/a considere más importantes.

Preguntas sobre el tema

El objetivo de *Principios y prácticas de evaluación del IB: una evaluación de calidad en la era digital* es aportar una visión general e integral del modo en que el IB aborda la evaluación; sin embargo, habrá quienes busquen información específica sobre alguna parte del proceso de evaluación. Con ese fin, las preguntas sobre el tema van seguidas de listas de secciones que se centran en preguntas concretas que los lectores/as pueden tener. Estas listas tienen carácter meramente informativo y no pretenden sugerir un orden de lectura de las secciones.

El glosario que figura al final de esta publicación también puede utilizarse para identificar términos de interés. Utilice la opción “Buscar” de la versión en PDF de esta publicación para obtener más información sobre cualquier término de interés del glosario.

“¿Cuál es el enfoque del IB acerca de la evaluación?”

Para el personal de admisión universitaria, docentes y partes implicadas que tienen interés en las bases teóricas de las evaluaciones del IB, la definición de evaluaciones de buena calidad según el IB, qué rige la toma de decisiones para establecer procesos, y cómo encontrar un equilibrio entre las demandas opuestas de la evaluación.

Ejemplos de preguntas habituales:

- ¿En qué consiste la evaluación del IB? ¿Qué la hace especial?
- ¿Qué cambios en la evaluación está causando la inteligencia artificial (IA)?

A continuación se indican las secciones que son particularmente pertinentes para estas preguntas.

Introducción y descripción general

- Buena evaluación
- Una buena evaluación sirve de apoyo a los objetivos curriculares
- Buena predictibilidad
- Una buena evaluación utiliza una variedad de tareas de evaluación
- La función de la evaluación realizada en clase y de la evaluación interna
- Trabajo colaborativo y puntuaciones individuales
- Una buena evaluación considera las competencias más amplias del alumnado y las habilidades de pensamiento de orden superior
- Habilidades de pensamiento de orden superior
- El perfil de la comunidad de aprendizaje y el desarrollo de habilidades
- Mentalidad internacional y entendimiento intercultural
- ¿Cómo es una buena evaluación digital?

Sección A: Principios de evaluación

- La evaluación en la educación
- Definición de la evaluación
- Enfoques de la evaluación
- El efecto de repercusión y el aprendizaje
- Validez
- Definición de la validez
- Creación de un argumento de validez
- Mantenimiento de la validez
- Elementos de la cadena de validez
- El equilibrio entre los distintos aspectos de la validez
- Fiabilidad
- Resultados coherentes y resultados “adecuados”
- Pertinencia del constructo y autenticidad
- Manejabilidad
- Equidad y sesgo
- Equiparabilidad
- El enfoque del IB acerca de la validez
- Beneficios de la evaluación digital para la validez
- Estándares
- Los tres aspectos de los estándares
- El enfoque normativo y el enfoque basado en criterios
- Enfoque normativo
- Enfoque basado en criterios
- Enfoque basado en logros
- Mantenimiento de los estándares
- Descripción del éxito del alumnado en la evaluación sumativa
- El impacto de las calificaciones
- La importancia del juicio profesional
- Corrección de las evaluaciones del IB
- Definición de corrección
- Métodos de corrección
- Corrección de la evaluación formativa

Sección B: Las prácticas de evaluación del IB

- Comunicación de los logros del alumnado
- El significado de las calificaciones finales del IB
- Diferencia entre puntuaciones y calificaciones
- Una convocatoria de exámenes satisfactoria
- Calificaciones y logros
- Definición de la integridad académica
- Priorización de la equidad para todo el alumnado

Recursos para imprimir

- “Los principios de evaluación del IB” (PDF)

“Soy profesional del PEP y en mi colegio no realizamos evaluaciones. ¿En qué me beneficia esta publicación?”

Para profesionales del PEP que deseen comprender las oportunidades de evaluación sumativa y de evaluación formativa que tienen los colegios.

Aunque el PEP no cuenta con evaluaciones que se envían al IB, en este programa existen oportunidades de evaluación sumativa y de evaluación formativa que se realizan en el colegio, y en esta publicación se describen buenas prácticas en esas áreas. Durante la primera infancia también es importante discutir las políticas relacionadas con la integridad académica y el fomento de una mentalidad ética, siempre de formas que resulten pertinentes para la edad del alumnado.

A continuación se indican las secciones que son particularmente pertinentes para estas preguntas.

Sección A: Principios de evaluación

- [Enfoques de la evaluación](#)
- [Corrección de la evaluación formativa](#)
- [Fomento de una mentalidad ética](#)
- [Uso ético de la inteligencia artificial en la evaluación](#)

Sección B: Las prácticas de evaluación del IB

- [Definición de la integridad académica](#)
- [Priorización de la equidad para todo el alumnado](#)

Sección C: Procesos específicos de los programas del IB

- [Elementos comunes a todos los programas](#)
- [Programa de la Escuela Primaria](#)

“Soy profesional del PAI y en mi colegio ya estamos realizando evaluaciones digitales. ¿En qué me beneficia esta publicación?”

Aunque hay elementos de esta publicación que tratan la transición a la evaluación digital en el PD y el POP, estas secciones también abordan el valor de las evaluaciones electrónicas del PAI. Por lo tanto, la publicación es pertinente para el PAI. Véase la pregunta sobre el tema “Soy profesional del PAI, el PD o el POP y me interesa saber cómo se corrigen las evaluaciones creadas por el IB y cómo se otorgan las calificaciones. ¿Qué comentarios hay disponibles para los colegios?”.

“¿En qué sentido la evaluación digital supondrá una diferencia para la evaluación del PD y el POP?”

Para el profesorado que está familiarizado con la evaluación del IB y que quiere saber cómo gestionará el IB el paso a las evaluaciones digitales.

Ejemplos de preguntas habituales:

- [¿En qué consiste la evaluación digital?](#)
- [¿El trabajo lo corregirá un dispositivo informático?](#)
- [¿Por qué será mejor?](#)
- [¿Se seguirán utilizando los mismos estándares?](#)

A continuación se indican las secciones que son particularmente pertinentes para estas preguntas.

Introducción y descripción general

- [Buena evaluación](#)
- [El papel de la tecnología en la evaluación del IB](#)
- [¿Cómo es una buena evaluación digital?](#)
- [Riesgos relacionados con la evaluación digital](#)

Sección B: Las prácticas de evaluación del IB

- Priorización de la equidad para todo el alumnado
- El ciclo de vida de la evaluación
- El impacto de la evaluación digital en el ciclo de evaluación

“Soy profesional del PAI, el PD o el POP y me interesa saber cómo se corrigen las evaluaciones creadas por el IB y cómo se otorgan las calificaciones. ¿Qué comentarios hay disponibles para los colegios?”

Para docentes que quieren entender qué sucede cuando envían los trabajos del alumnado al IB para la corrección y cómo se conceden las calificaciones. Es posible que también quieran entender por qué el IB incluye la evaluación del profesorado en sus procesos, pero posteriormente no siempre acepta las valoraciones que hace el profesorado.

Ejemplos de preguntas habituales:

- ¿Por qué este año los límites de calificación son distintos?
- ¿Por qué no puedo tener otro examinador/a?
- ¿Quién corrige los exámenes?
- ¿Cómo se comprueba el trabajo de los examinadores/as?
- ¿Cómo puedo presentar una apelación?
- Yo le hubiera concedido una puntuación diferente al alumno/a; ¿por qué es esto justo?
- El esquema de calificación indica que la respuesta del alumno/a merece una puntuación de X; ¿cómo puede ser esto justo?
- ¿No sería mejor para el alumnado tener un “aula sin calificaciones”?
- ¿Por qué han cambiado mis puntuaciones?
- ¿Por qué mi factor de moderación cambia de un año a otro?

A continuación se indican las secciones que son particularmente pertinentes para estas preguntas.

Introducción y descripción general

- La función de la evaluación realizada en clase y de la evaluación interna

Sección A: Principios de evaluación

- La evaluación en la educación
- Definición de la evaluación
- Enfoques de la evaluación
- El efecto de repercusión y el aprendizaje
- Elementos de la cadena de validez
- El impacto de las calificaciones
- Corrección de las evaluaciones del IB

Sección B: Las prácticas de evaluación del IB

- Comunicación de los logros del alumnado
- El significado de las calificaciones finales del IB
- Diferencia entre puntuaciones y calificaciones
- Una convocatoria de exámenes satisfactoria
- Calificaciones y logros
- Calificaciones previstas
- El proceso de evaluación: funciones y responsabilidades
- Examinador principal y examinador jefe
- Otros examinadores
- Las responsabilidades del personal del IB

- Jerarquía del personal examinador
- Priorización de la equidad para todo el alumnado
- Puntuación y calificación
- Enfoques de la corrección
- Esquemas de calificación analíticos
- Criterios holísticos: bandas de puntuación
- Estandarización
- El modelo de calidad
- Respuestas de práctica
- Respuestas de aptitud
- Corrección en condiciones reales y respuestas de control
- Indicadores de una estandarización satisfactoria
- Márgenes de tolerancia
- Respuestas complicadas o inusuales
- Vínculos con los colegios
- Grupos de preguntas
- Totalización
- Moderación
- Moderación y expectativas del profesorado
- Selección de trabajos del alumnado
- Situaciones irregulares
- Casos en los que no es posible encontrar un factor de moderación
- Toma dinámica de muestras
- Concesión de calificaciones finales y totalización
- Límites de calificación estimativos e interpolados
- Pruebas utilizadas en el proceso de concesión de calificaciones finales
- Consideración del grupo de estudiantes del año en curso
- Comentarios y opiniones sobre la evaluación
- Búsqueda de pruebas en los exámenes
- Análisis de las estadísticas sobre los resultados
- Cómo encontrar el equilibrio entre las pruebas
- Límites de calificación fijos
- Revisión y aprobación de las recomendaciones sobre los límites de calificación
- Otorgamiento de la titulación de un programa
- Profesorado observador
- Principios del proceso de concesión de calificaciones finales
- Controles de calidad
- Comentarios para los colegios
- Informes sobre la evaluación interna

Sección C: Procesos específicos de los programas del IB

- Elementos comunes a todos los programas

“¿Cómo diseña y desarrolla el IB los exámenes?”

Para docentes que quieren entender cómo se preparan los cuestionarios de examen, y para aquellas personas a las que le preocupa la calidad de dichos cuestionarios

Ejemplos de preguntas habituales:

- ¿Quién redacta los exámenes?
- ¿Cómo se verifican?
- ¿Su contenido es el mismo cuando están en lenguas distintas?
- ¿Realmente evalúan lo correcto?

A continuación se indican las secciones que son particularmente pertinentes para estas preguntas.

Introducción y descripción general

- Buena evaluación
- Una buena evaluación sirve de apoyo a los objetivos curriculares
- Buena predictibilidad
- Una buena evaluación utiliza una variedad de tareas de evaluación
- La función de la evaluación realizada en clase y de la evaluación interna

Sección A: Principios de evaluación

- La evaluación en la educación
- Definición de la evaluación
- Enfoques de la evaluación
- El efecto de repercusión y el aprendizaje
- Validez
- Definición de la validez

Sección B: Las prácticas de evaluación del IB

- Funciones relacionadas con la creación de exámenes
- Priorización de la equidad para todo el alumnado
- Elaboración de evaluaciones
- Descripción general de la producción de exámenes
- El proceso de aprobación del contenido
- Creación y aprobación del contenido
- Formato y revisión
- Aprobación de la usabilidad
- Control de calidad
- Traducciones
- Pruebas de examen modificadas
- Puntuación y calificación

Principios y prácticas de evaluación

Los **principios de evaluación** son las propuestas de base que se consideran importantes para crear e implementar programas y evaluaciones, así como al corregir y otorgar calificaciones. Estos principios proceden de los elementos esenciales de la educación del IB. El más importante es que las evaluaciones deben servir para respaldar la educación, no para distorsionarla.

Las **prácticas de evaluación** son las maneras en las que el IB aplica sus principios de manera concreta y significativa. Dichas prácticas tienen en cuenta las demandas opuestas y las limitaciones prácticas de trabajar a escala mundial, y mantienen al mismo tiempo la integridad, que es uno de los valores clave de la filosofía del IB.

A continuación se resumen los principios de evaluación del IB.

La evaluación, ya sea sumativa o formativa, debe:

- Resultar válida para el propósito para el que está pensada. Esto significa que debe haber un equilibrio entre las demandas opuestas de la pertinencia del constructo (es decir, evaluar lo correcto), la fiabilidad, la equidad (es decir, que no haya sesgos), la equiparabilidad con otras evaluaciones, y la manejabilidad para estudiantes, colegios y el IB.
- Tener un efecto de repercusión positivo: su diseño debe favorecer un aprendizaje y una enseñanza de buena calidad.
- Resultar adecuada para el mayor número de estudiantes posible, de manera que puedan demostrar su nivel de logro personal.
- Estar diseñada de manera que se puedan realizar todas las adecuaciones necesarias para garantizar el acceso para todo el alumnado.
- Formar parte integrante del programa del IB, y no considerarse como un elemento aislado.
- Favorecer la simultaneidad del aprendizaje y la experiencia del alumnado en general.
- Respalda los principios del IB y formar personas que tengan los atributos de indagación, conocimiento, razonamiento, comunicación y mentalidad abierta.

Las prácticas de evaluación se describen de forma general y se deben aplicar adecuadamente dentro del contexto de asignaturas, disciplinas y programas individuales. Las prácticas de evaluación explican de manera amplia qué hay que hacer, pero sin entrar en detalles sobre los procesos cotidianos y las diferencias para aplicarlas en los distintos programas y asignaturas.

El recurso para imprimir "[Los principios de evaluación del IB](#)" (PDF) está disponible para el profesorado.

Buena evaluación

- No existe una definición única de *buena evaluación*; todo depende de la importancia relativa que se dé a las distintas prioridades, y del propósito de la evaluación en cuestión.
- En el IB, el principio subyacente es que hay que evaluar aquello que es importante, y no considerar automáticamente importante aquello que se puede evaluar.
- La evaluación del IB busca tener un efecto de repercusión positivo en el aprendizaje y la enseñanza.
- Por regla general, las buenas evaluaciones deben incluir una variedad de tareas de desempeño que ayuden a mostrar un perfil completo del aprendizaje, y permitir organizar actividades de mayor profundidad en clase, además de exámenes.
- Para incluir a todo el alumnado con diferentes preferencias y estilos de aprendizaje, las buenas evaluaciones deben diseñarse y desarrollarse teniendo en cuenta los principios del diseño universal de evaluación. Dichas evaluaciones deben considerar múltiples formas de participación, representación, acción y expresión.

Desde el punto de vista del IB, una **buena evaluación da prioridad a los resultados de aprendizaje significativos**. Asimismo, debe tener una fuerte correspondencia con los objetivos curriculares y ayudar al alumnado a desarrollar una amplia variedad de habilidades y competencias.

El principio subyacente del IB es que hay que **evaluar aquello que es importante**, y no considerar automáticamente importante aquello que se puede evaluar. Además, debe haber un equilibrio entre esto y otras consideraciones, como la fiabilidad y la carga de trabajo del alumnado.

Es difícil que un solo enfoque pueda satisfacer todas las prioridades. Específicamente, un buen diseño de evaluación será distinto para la evaluación sumativa y la evaluación formativa. Partiendo de este principio, los aspectos que el IB vincula con una buena evaluación son:

- Apoyo a los objetivos curriculares
- Uso de una variedad de tareas de evaluación
- Consideración de las competencias más amplias del alumnado y de las habilidades de pensamiento de orden superior

En las siguientes secciones se abordan uno por uno estos aspectos.

Una buena evaluación sirve de apoyo a los objetivos curriculares

- Las evaluaciones deben tener un efecto de repercusión positivo que fomente un buen aprendizaje y una buena enseñanza.
- Las evaluaciones deben tener una buena predictibilidad.

La evaluación no se debe considerar como una actividad separada del aprendizaje y la enseñanza. Los resultados de las evaluaciones del IB se basan en la evaluación sumativa y no pretenden proporcionar comentarios directos sobre el aprendizaje y la enseñanza. Sin embargo, está claro que lo que se incluye en la evaluación tendrá un impacto en lo que se enseña. Esto se conoce como *efecto de repercusión*.

Para el IB, el diseño de las evaluaciones debe fomentar los resultados educativos más deseados para el alumnado. Los efectos sobre el aprendizaje del alumnado continúan siendo una consideración esencial en

el diseño de las evaluaciones del IB y, junto con la pertinencia del constructo, ayudan a orientar la forma en que el IB encuentra el equilibrio entre los distintos elementos de la validez.

La evaluación de alto impacto tiene una considerable influencia sobre el aprendizaje y la enseñanza. Para aprovechar esto, se pueden diseñar evaluaciones que fomenten una buena pedagogía y la implicación constructiva del alumnado en su propio aprendizaje (véase, por ejemplo, Murphy, 1999).

En los programas del IB, la evaluación está estrechamente ligada a la experiencia de aprendizaje y es un reflejo de ella. Mediante el uso del enfoque de correspondencia constructiva (Biggs, 1996), todo se diseña en torno a los resultados del aprendizaje esperados. Estos resultados influyen y orientan tanto la forma en que se enseña al alumnado como la forma en que se lo evalúa, a fin de que cuente con una buena preparación para obtener buenos resultados. Para obtener más información, consulte la publicación *Enfoques de la evaluación en el IB*.

Las características personales que se desea que tenga el alumnado, expresadas en la declaración de principios del IB y exploradas en el perfil de la comunidad de aprendizaje del IB, se ajustan muy bien a la **teoría constructivista del aprendizaje**. En dicho marco, el alumnado se implica activamente en el proceso de aprendizaje, se responsabiliza de su propio aprendizaje y amplía sus conocimientos, comprensión y destrezas a través de la indagación.

Por ejemplo, en los requisitos de evaluación de varias asignaturas, se espera que el alumnado participe de manera activa y abierta en perspectivas culturales distintas de la suya. Las cualidades de naturaleza afectiva, como son la solidaridad y la compasión, resultan difíciles de incluir en la evaluación formal, pero deben, no obstante, estar representadas en el sistema de evaluación global. Esto se consigue parcialmente mediante requisitos específicos del currículo de cada programa que no se evalúan, como la exposición en el PEP, el elemento de Compromiso Comunitario en el PAI, el requisito de Creatividad, Actividad y Servicio (CAS) en el PD, y Compromiso Comunitario en el POP.

En lo que se refiere al diseño de los cursos del PD y el POP en particular, el IB pone mucho énfasis en la validez predictiva (la medida en que los resultados predicen el éxito futuro), consciente de que el modo en que se lleve a cabo la evaluación tendrá una influencia decisiva en la forma en que se enseñen los cursos del IB en los colegios. El modelo de evaluación (conjunto de instrumentos de evaluación) aplicado a cada asignatura se diseña para que tenga una base amplia que abarque diversos tipos de datos. El objetivo es fomentar la pertinencia del constructo al proporcionar la mayor variedad posible de pruebas que respalden los logros y el aprendizaje del alumnado.

Aunque el IB es muy consciente de que las evaluaciones tienen un efecto de repercusión en el aprendizaje y la enseñanza, también insta a los colegios a adoptar pedagogías que desarrollen en el alumnado todos los objetivos y filosofías de los programas del IB.

El IB realiza con regularidad investigaciones para evaluar en qué medida los programas diseñados sirven como buena preparación para estudios posteriores. Para obtener más información sobre los estudios que realiza el IB, consulte la sección *"Investigación"* del sitio web público del IB.

Buena predictibilidad

El término *predictibilidad* designa la capacidad de estimar qué sucederá o cuándo sucederá algo. En el ámbito de la evaluación, la predictibilidad se refiere a la capacidad de los colegios para determinar qué tipos de preguntas aparecerán en un examen determinado y cuándo. Una buena predictibilidad en la evaluación contribuye a la equidad, ya que permite que las preguntas concuerden con el currículo que el profesorado utiliza para preparar a sus estudiantes. Esto garantiza que el alumnado sepa qué tipos generales de preguntas puede esperar, aunque el contenido exacto varíe.

El IB quiere asegurarse de que la inversión de los colegios en sus opciones de enseñanza se vea recompensada durante el tiempo que dura un currículo concreto (antes de que se revise). Todo el programa de estudios se debe examinar del modo en que lo requiere la evaluación específica. Hacia el final de un curso, se presta atención para eliminar el riesgo de mala predictibilidad (cuando el colegio identifica lo que todavía no se ha preguntado y, por lo tanto, es probable que aparezca en un examen).

El principio subyacente es que nada de lo que aparezca en un examen determinado debería ser una sorpresa, ni para el alumnado ni para el colegio. Por tanto, sea cual sea el componente, las preguntas

planteadas deben basarse explícitamente en la guía de la asignatura que corresponda. Cuando es posible y resulta adecuado en las evaluaciones, el IB procura eliminar los problemas causados por la predictibilidad. Para ello, diseña las evaluaciones de tal modo que no se favorezca a aquellos alumnos/as que lleguen con respuestas preparadas con antelación.

El diseño de la evaluación resulta primordial para que haya suficientes maneras de evaluar todo tipo de tema, opción o texto, con el fin de mitigar la mala predictibilidad.

Tabla 3
Buena y mala predictibilidad

Buena predictibilidad	Mala predictibilidad
Si se elaborara una lista de todas las preguntas que se podrían plantear sobre un tema, una opción o un texto en función de los términos de instrucción, aparecerían las preguntas formuladas por el IB para esa asignatura.	El profesorado prepara en exceso a sus estudiantes para determinadas preguntas cuando detectan que varían muy poco de una convocatoria de exámenes a otra. Esto resulta problemático para las preguntas de respuestas más largas, que requieren demostrar múltiples habilidades (conocimiento, análisis y evaluación).
Las preguntas se reutilizan de manera impredecible.	Las preguntas posibles están limitadas debido a malas decisiones tomadas durante el diseño de la evaluación (por ejemplo, establecer para un curso un texto prescrito que solo tiene un tema principal sobre el que se pueden hacer preguntas). Esto nos lleva a una mala predictibilidad debido al diseño.

Una buena evaluación utiliza una variedad de tareas de evaluación

Una pregunta de opción múltiple, una pregunta de respuesta corta, una pregunta de respuesta extensa, un ensayo, un proyecto, un trabajo de carpeta o un trabajo de investigación son ejemplos de tareas de evaluación.

Un instrumento o componente de evaluación consta de una o más tareas que se reúnen por motivos de continuidad temática o de contenido, o por conveniencia. Un examen sumativo, una carpeta, un proyecto de creación colectiva o un trabajo de investigación son ejemplos de instrumentos o componentes de evaluación. Los conceptos de *tarea de evaluación* y de *componente* se solapan. A veces, un alumno/a puede realizar una sola tarea de entre varias a elegir para un componente.

Existen varios motivos por los que el IB utiliza una gran diversidad de tareas y componentes de evaluación. En primer lugar, desde una perspectiva histórica y pragmática, Peterson (2003) dice respecto al desarrollo inicial de la evaluación en el PD que “teníamos tanto una obligación como una oportunidad de tener en cuenta las diferentes técnicas de evaluación utilizadas en aquellos países y regiones o territorios a cuyas instituciones la mayoría del alumnado del IB pretendía acceder”, y este principio se extiende al PAI y al POP. Existen también consideraciones de validez, relativas a la adecuación al propósito, que exigen un enfoque variado de la evaluación. Por último, la variedad de técnicas de evaluación contribuye a reducir la posibilidad de falta de equidad en la evaluación, como indican Linn (1992) y Brown (2002). La gama de componentes del modelo de evaluación de una asignatura y el conjunto de tareas inherentes garantizan una representación adecuada de los logros del alumno/a con respecto a los objetivos específicos de esa asignatura.

La función de la evaluación realizada en clase y de la evaluación interna

Las evaluaciones realizadas en clase ofrecen múltiples oportunidades para poner a prueba al alumnado en áreas que no resultan adecuadas para los exámenes. Su principal ventaja es que permiten pedirles que

Lleven a cabo una tarea extensa con la que puedan investigar un problema y mostrar cómo desarrollan su pensamiento sin tener que someterse a las restricciones de tiempo características de los exámenes. Esto significa que hay una amplia variedad de tareas de evaluación que solo se pueden aplicar mediante una evaluación realizada en clase.

Entre las tareas que se prestan a una evaluación realizada en clase se encuentran los proyectos, los trabajos de campo, los trabajos prácticos de laboratorio, las representaciones teatrales y las investigaciones matemáticas.

El trabajo evaluado internamente tiene otras **ventajas** en el marco de las titulaciones internacionales. Dicho tipo de evaluación permite mayor **flexibilidad** en la elección de los temas y, a la vez, seguir abordando un conjunto de habilidades comunes. Los colegios pueden así situar el estudio en un **contexto local, cultural o geográfico**, o establecer lazos más estrechos entre el aula y el mundo exterior. Los colegios internacionales, cuyo alumnado a menudo procede de una cultura distinta a la del lugar donde se ubica el colegio, pueden utilizar el trabajo de evaluación interna para implicarse de forma más estrecha en la sociedad o el entorno locales.

Además, la evaluación interna a menudo ofrece a cada estudiante la oportunidad de elegir sus propios temas o cuestiones. Esto les permite seguir sus intereses particulares y tener mayor control de su propio aprendizaje. Esta flexibilidad de enfoque hace que la evaluación interna sea una valiosa adición a la educación del alumnado, ya que incrementa la validez del proceso de evaluación y de la experiencia de aprendizaje en su conjunto.

Sin embargo, también surgen algunos **desafíos** en la evaluación realizada en clase. Uno de los principales es garantizar la **integridad académica**. Cuando las tareas no se completan bajo supervisión estándar (por ejemplo, si se realizan fuera del aula), es mucho más difícil garantizar que el alumnado no contravenga la política de integridad académica del IB (por ejemplo, pedir a otra persona que le haga el trabajo). Con libre acceso a Internet, este desafío aumenta de manera exponencial. Por eso, el IB considera que el personal del colegio y el profesorado son las personas más indicadas para ayudar a verificar la autoría original de los trabajos, y para identificar los casos en los que el alumnado no entrega su propio trabajo. Para obtener más información sobre la integridad académica, y recursos al respecto, consulte la página [“Integridad académica”](#) del sitio web público del IB.

Otra dificultad es que la evaluación realizada en clase puede suponer una **carga de trabajo** importante tanto para docentes como para estudiantes. Al realizar la evaluación en clase se reducen las horas lectivas disponibles, y las tareas que se hacen internamente son normalmente sustanciales y exigen que el alumnado les dedique mucho tiempo. Aunque es correcto que el profesorado emplee tiempo en preparar a sus estudiantes en las habilidades y procesos necesarios para la evaluación interna, puede existir una fuerte tentación, por parte de estudiante y docente, de ensayar y practicar la tarea concreta establecida para la evaluación interna más de lo necesario. Esto implica reducir aún más el tiempo dedicado a la enseñanza. En las guías de las asignaturas del PD se ofrece orientación detallada sobre el papel del profesorado al prestar apoyo en los trabajos de evaluación interna.

La evaluación realizada en clase se puede corregir **externamente** o **internamente**. Cuando se corrige externamente, el trabajo que hace el alumno/a se envía al IB y lo corrige un examinador/a. El IB supervisa la calidad de las decisiones del examinador/a (véase la sección [“Corrección”](#) de esta publicación).

En cambio, en el caso de la evaluación interna, el colegio (normalmente el profesor/a) corrige el trabajo del alumno/a. Posteriormente, el IB, mediante un proceso de moderación, comprueba que el profesor/a haya aplicado correctamente el estándar global (véase la sección [“Moderación”](#) de esta publicación).

Existen diversas visiones sobre la evaluación interna en todo el mundo. Algunos sistemas educativos conceden mucha importancia al hecho de que el profesorado es quien se encuentra en la mejor posición para dar una opinión holística acerca del desempeño del alumnado, evitando así las limitaciones inherentes de realizar una sola evaluación. Por su parte, en otros sistemas educativos, el desempeño del profesorado se determina a partir de los resultados del alumnado, lo cual supone un gran incentivo para que concedan puntuaciones más altas en todas las evaluaciones internas.

Asimismo, la opinión del profesor/a puede verse influida si conoce trabajos previos del alumno/a, lo cual puede introducir diversos tipos de sesgos. Puede que el profesorado a veces no tenga claros los límites de su papel al orientar y apoyar a sus estudiantes mientras llevan a cabo el trabajo de evaluación interna, y es

posible que tenga solo una visión limitada de los estándares globales de logro que hay en su área disciplinaria. Cuando evalúa el trabajo de sus estudiantes, el profesorado puede verse muy influenciado por los estándares generales existentes en su colegio Incluso cuando no hay ningún incentivo para conceder a los trabajos entregados para evaluación puntuaciones más altas de lo que correspondería, la relación profesional que el profesorado establece con el alumnado provoca que sea difícil tomar decisiones objetivas. Los estudios realizados sobre este sesgo inconsciente indican que puede tener un efecto notable (véase, por ejemplo, Zanga y De Gioannis, 2023).

Para el IB, los beneficios de utilizar tareas de evaluación significativas que incluyan trabajos realizados en clase, corregidos por alguien que haya sido testigo del desarrollo del alumnado, superan los riesgos que implica la evaluación interna. Por ese motivo, normalmente las tareas de evaluación interna forman parte fundamental del modelo de evaluación de una asignatura.

Trabajo colaborativo y puntuaciones individuales

Una cuestión complicada es cómo evaluar las tareas colaborativas, en las que se espera que el alumnado realice un trabajo colectivo para completar una actividad. Este es un aspecto importante que el alumnado encontrará en el mundo laboral, pero que se obvia a menudo en la evaluación académica.

En algunas circunstancias, es posible identificar el papel de cada persona en la actividad en grupo. Un ejemplo de esto podría ser una representación de Danza en la que se puedan ver las habilidades de cada bailarín/a y así se le pueda conceder una puntuación individual. En estas situaciones, el trabajo en grupo no suscita controversia.

En otros casos, no es posible identificar en qué grado ha contribuido cada persona al trabajo final. En dichos casos, es posible conceder una única puntuación para todo el grupo, pero esto implica que no se tienen en cuenta las diferencias de contribución de cada estudiante al resultado total. Es posible que un alumno/a se haya encargado de la mayor parte del trabajo y haya aportado la mayoría de los conocimientos, pero, a pesar de todo, no recibiría mayor reconocimiento que el resto. Dado que los resultados del IB se utilizan para tomar decisiones de selección, se considera que este modo de proceder no sería justo. Por este motivo, en general las evaluaciones en grupo se deben evitar si no permiten medir los logros de cada estudiante individual.

Una buena evaluación considera las competencias más amplias del alumnado y las habilidades de pensamiento de orden superior

La educación del IB aspira a lograr mucho más que el simple aprendizaje de “hechos” por parte del alumnado. Este arraigado objetivo del IB se refleja actualmente en **iniciativas** como las que adoptan diversos gobiernos para dotar al alumnado de **habilidades** del siglo XXI y **competencias laborales**.

La filosofía del IB y su enfoque acerca de las competencias del alumnado se centran en el **perfil de la comunidad de aprendizaje del IB** y su relación con la **mentalidad internacional**. En la “Sección C: Procesos específicos de los programas del IB”, que aborda la continuidad entre los programas, se aporta más información al respecto. El concepto más importante que hay que tener en cuenta es que el enfoque del IB acerca de la evaluación garantiza que se evalúe a cada estudiante no solo sobre sus conocimientos de los contenidos, sino también sobre diversas competencias que tienen aplicaciones en el mundo real. Mediante un diseño cuidadoso, las evaluaciones del IB tienen como objetivo formar **personas que adopten una actitud de aprendizaje durante toda su vida** y que estén bien preparadas para futuros desafíos.

Por lo tanto, una buena evaluación tiene en cuenta **toda la variedad de resultados** que cada curso pretende lograr, y permite que el alumnado demuestre sus habilidades en dichos resultados. Sin embargo, con frecuencia solo resulta deseable o factible medir una pequeña parte de los resultados, y una evaluación de buena calidad también equilibra estas limitaciones. Para el IB, el mejor sistema consiste en categorizar estos resultados en función de las **habilidades de pensamiento de orden superior**, de las **competencias** más amplias del alumnado, y de su relación con la **mentalidad internacional**.

Habilidades de pensamiento de orden superior

Las evaluaciones del IB están diseñadas para medir las habilidades de pensamiento de orden superior que son la capacidad de **evaluación** y la capacidad de **análisis**, y no la simple facultad de memorizar conocimientos.

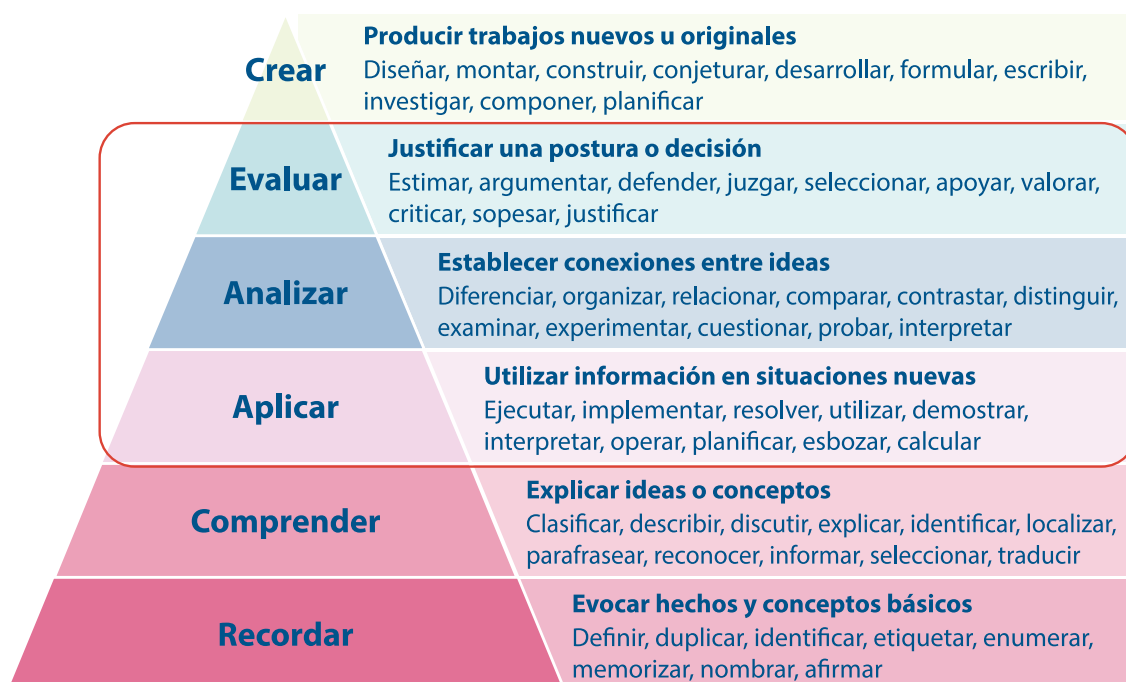
Así lo señaló Alec Peterson (2003), cuyas opiniones dieron forma a la filosofía educativa del IB. En ese sentido, afirmó que “lo que importa no es la absorción y regurgitación de hechos o de interpretaciones predigeridas de hechos, sino el desarrollo de las facultades de la mente o de formas de pensar que pueden aplicarse a nuevas situaciones y a la presentación de nuevos hechos según van surgiendo”. Sugata Mitra, en su intervención en la Conferencia para directores de Colegios del Mundo del IB celebrada en Singapur en 2011, llevó la argumentación de Peterson un paso más allá y sostuvo que la disponibilidad inmediata e integral de conocimientos a través de Internet significa que los conocimientos en sí tienen poco valor. Más bien, sostiene Mitra, lo que necesitan las personas del siglo XXI es la capacidad de analizar, interpretar y seleccionar conocimientos (Mitra, 2011).

En las evaluaciones del IB se presta mucha atención a las habilidades de pensamiento de orden superior. Con este fin, por ejemplo, las expectativas de desempeño en las tareas de evaluación se basan en taxonomías establecidas, como la taxonomía de Bloom (Bloom *et al.*, 1956; Anderson y Krathwohl, 2001). De hecho, la taxonomía de objetivos educativos de Bloom proporciona un marco útil para expresar la diversidad de habilidades y procesos cognitivos necesarios para completar satisfactoriamente las tareas de evaluación y obtener las máximas puntuaciones. Las capacidades cognitivas de orden superior a las que se refiere Bloom exigen ciertamente que se utilice un tipo distinto de evaluación. Las capacidades de análisis, síntesis y evaluación del alumnado solo pueden valorarse adecuadamente si se le exige analizar, sintetizar y evaluar con cierta amplitud. La evaluación del desempeño es el único modo realista de medir los logros en estas áreas y, puesto que los resultados de tal actividad no pueden prescribirse de manera estricta, estas evaluaciones deben ser relativamente desestructuradas y abiertas, de tal modo que puedan existir muchas respuestas distintas pero correctas.

Figura 2

Taxonomía de Bloom

¿Es el alumno capaz de lo siguiente?



Para que sean pertinentes, las evaluaciones del IB deben ser capaces de reconocer y recompensar el desempeño del alumnado en estas habilidades, a pesar de las dificultades que presentan en relación a la

fiabilidad y otros aspectos de la validez. Las pruebas que solo premian la capacidad de memorizar conocimientos, conceptos y técnicas cotidianas no resultan adecuadas al propósito de la educación del IB, ya que no abordan de manera adecuada estas importantes habilidades de pensamiento de orden superior.

El perfil de la comunidad de aprendizaje y el desarrollo de habilidades

Actualmente, la educación trata mucho más sobre formas de pensar que implican enfoques creativos y críticos para la resolución de problemas y la toma de decisiones. Asimismo, aborda métodos de trabajo, en particular de comunicación y colaboración, así como las herramientas necesarias, tales como la capacidad de reconocer y explotar el potencial de las nuevas tecnologías, y también de evitar sus riesgos. Por último, la educación abarca la capacidad de vivir como ciudadano activo y comprometido en un mundo polifacético. Estos ciudadanos influyen en lo que quieren aprender y cómo quieren hacerlo, y eso da forma al papel de los educadores.

(Schleicher, 2016)

Cada vez se afirma más que el siglo XXI requiere habilidades radicalmente distintas a las que se esperaban en las generaciones anteriores, aunque hay quienes argumentan que el enfoque de indagación que subyace a las habilidades del siglo XXI se ha valorado desde Sócrates. Sin embargo, existe cierto consenso general sobre la importancia de proporcionar al alumnado una amplia variedad de atributos que sirva de preparación para la vida. Véanse, por ejemplo, los argumentos que da Llewellyn (2014).

Hay muchas maneras distintas de categorizar estas habilidades, por ejemplo, el marco de competencias para el siglo XXI de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), el que propone RAND Education, y el del National Research Council (NRC). En el IB, estas competencias se definen y se describen en el perfil de la comunidad de aprendizaje.

No todos los aspectos del perfil de la comunidad de aprendizaje se pueden medir adecuadamente con la evaluación sumativa, pero son varios los que quedan englobados en el concepto de habilidades de pensamiento de orden superior. Una buena evaluación reconoce la importancia de dichas características y, a través de oportunidades formativas y sumativas, crea un espacio para el desarrollo de estas competencias. Algunos ejemplos serían fomentar enfoques éticos para las encuestas y los experimentos en Psicología y Ciencias, la exploración amplia de un dilema ético en el Proyecto de Reflexión, respaldar una revisión adecuada entre compañeros/as, e introducir contextos inesperados para el alumnado.

Para obtener más información sobre el enfoque amplio del IB respecto a las competencias del alumnado, consulte el material disponible en el [sitio web del IB](#) o la publicación *De los principios a la práctica* del programa pertinente.

Mentalidad internacional y entendimiento intercultural

El IB tiene como meta ayudar al alumnado a convertirse en “personas solidarias [...] capaces de contribuir a crear un mundo mejor y más pacífico, en el marco del entendimiento mutuo y el respeto intercultural” y a alentarle a “entender que otras personas, con sus diferencias, también pueden estar en lo cierto” (Declaración de principios del IB, 2024). El alumnado que estudia los programas del IB se encuentra en numerosos países y regiones o territorios, y tiene una amplia variedad de nacionalidades. Por lo tanto, en el aprendizaje y la enseñanza del IB hay un **contexto internacional** y un propósito de desarrollar una **mentalidad internacional**, y ambos elementos deben reflejarse en la evaluación. Un paso importante para lograr esta diversidad de perspectivas es emplear a especialistas en el ámbito académico —entre otros, creadores/as de contenidos de evaluación y responsables de desarrollo curricular— procedentes de distintos contextos culturales. Para cumplir los principios del IB es importante no ocultar las diferencias, sino utilizarlas de manera que el alumnado pueda explorarlas de manera productiva y positiva.

En algunas áreas disciplinarias, para fomentar la diversidad cultural se pueden tener en cuenta distintas perspectivas culturales y hacer hincapié en ofrecer un contenido culturalmente variado en el currículo. Podemos encontrar ejemplos de este enfoque en varias asignaturas del PD, como Biología, Química, Psicología y Artes Visuales. En las tres primeras, las estructuras opcionales permiten a los colegios seleccionar un contenido que, en cierta medida, encajará con sus tradiciones culturales acerca de la enseñanza de la asignatura.

En otras áreas disciplinarias, como Artes y Lengua y Literatura, la mentalidad internacional se potencia a través del contenido curricular recomendado y prescrito. También se puede hacer hincapié en esto mediante un conjunto más amplio de tareas de evaluación interna.

La mentalidad internacional no se reduce a conocer y comprender otras culturas. Las **actitudes** que se muestran y las **acciones** que se realizan también son atributos importantes. Las actitudes son difíciles de juzgar mediante la evaluación escolar tradicional.

En el IB esto se aborda con los componentes troncales que no se evalúan de sus programas, como el Proyecto Comunitario en el PAI y el componente de CAS en el PD. Como el alumnado solo puede obtener el certificado del PAI, el diploma del IB y el certificado del POP si completa los componentes troncales del programa, estos elementos tienen un impacto significativo en el resultado global de la evaluación del IB.

Figura 3

Acción basada en principios



Tal vez se considere que permitir al alumnado elegir qué preguntas quiere responder es la mejor manera de abordar los distintos requisitos internacionales en materia de evaluación. Sin embargo, este enfoque provoca problemas en lo que se refiere a mantener la **equiparabilidad** entre las opciones. Esto sucede con mayor frecuencia cuando se pueden elegir las preguntas, o cuando las tareas de evaluación son muy abiertas, lo que hace aún más complicado definir en qué consiste una “exigencia igual” cuando el alumnado tiene bagajes educativos muy distintos. En general, la manera más fácil de mantener la equiparabilidad es establecer tareas comunes que permitan al alumnado apoyarse en sus propias experiencias para dar las respuestas. En esos casos, el personal examinador afronta el reto de mantener un estándar común.

Para obtener información sobre la equiparabilidad, se puede analizar el desempeño del alumnado. Este análisis se trata con más profundidad en la sección “[Concesión de calificaciones finales y totalización](#)”.

La evaluación que se lleva a cabo en un contexto internacional plantea más retos en cuanto a la equidad de los que normalmente se encuentran en el seno de un sistema nacional. Preguntas que pueden ser perfectamente adecuadas en un marco nacional se convierten en inadecuadas en otro. Por ejemplo, las preguntas que se refieren a deportes, viajes, entretenimiento, acontecimientos históricos, incluso al tiempo meteorológico, deben prepararse y tratarse con sumo cuidado. Puede parecer que el único modo de evitar

este problema sea preparar preguntas de examen carentes de todo contexto sociocultural. Sin embargo, hacerlo así no solo limitaría mucho las preguntas de exámenes, sino que también contradeciría la filosofía de evaluación del IB y el correcto ejercicio de la actividad de evaluación en general, que procura garantizar su validez mediante tareas basadas en contextos. La evaluación y el trabajo contextualizados son esenciales para el buen aprendizaje.

Existen dos posibles formas de abordar este dilema. El primer enfoque consiste en proporcionar al alumnado información contextual adecuada. Para ello, se pueden aportar especificaciones en el programa de estudios de la asignatura, proporcionar estudios de caso en los que se basen las preguntas, o incluso incorporar dicha información contextual en el contenido del examen.

El segundo enfoque consiste en utilizar preguntas y tareas de evaluación más abiertas que permitan al alumnado utilizar su propio contexto para crear sus respuestas. Es importante tener en cuenta que, en este enfoque, la corrección ha de centrarse en niveles de comprensión más profundos, y no en conocimientos directos del contenido de la asignatura, ya que puede que no haya una base común de contenido. Esto es coherente con la filosofía de evaluación del IB.

Figura 4

Variedad de contextos culturales



Incluso aplicando estos dos métodos, el alumnado puede encontrarse con tareas de evaluación cuyos contextos socioculturales no le resulten familiares. También esto es coherente con la filosofía de evaluación del IB, dado que uno de sus objetivos es animar al alumnado a tener una **mentalidad más abierta**, una **mayor conciencia global**, y mayor **comodidad** y **competencia** al desenvolverse en contextos culturales que no le sean familiares.

Un número significativo de estudiantes del IB se presentan a exámenes en una lengua que no es su primera lengua ni aquella en la que son más competentes. En casi todos estos casos se trata del inglés, porque el alumnado que estudia en francés o español (las otras dos lenguas en las que se realiza la evaluación con mayor frecuencia) suele hablarlas con fluidez. Por consiguiente, durante la creación de las evaluaciones y las revisiones editoriales, se presta especial atención a que la redacción de las preguntas no perjudique al alumnado que realiza la evaluación en las lenguas adicionales. Esta es una prioridad importante durante el proceso de edición de los exámenes.

El papel de la tecnología en la evaluación del IB

- El alumnado utiliza la tecnología en muchas situaciones, como en su vida social, para el aprendizaje, y para escribir sus trabajos. Además, continuarán utilizando tecnología ampliamente en sus estudios posteriores y en su vida profesional. Por lo tanto, el uso de la tecnología en los exámenes es coherente con la experiencia que tienen del mundo.
- El IB utiliza la tecnología para ofrecer una evaluación de mayor calidad y significativa. El uso de la tecnología no influye en los enfoques de la evaluación que se eligen.
- Es importante distinguir entre el efecto de la tecnología para respaldar al personal examinador experto (corrección electrónica) y el uso de la tecnología para crear evaluaciones significativas para el alumnado.
- El alumnado deberá aprender en clase a utilizar la tecnología que se utilizará para realizar la evaluación.

Evaluación digital

La evaluación digital es aquella que se realiza en un dispositivo informático. Esto puede implicar muchos estilos diferentes de preguntas y estímulos, según lo que se evalúe: desde evaluaciones con contenidos multimedia y muy interactivas, hasta exámenes más tradicionales con preguntas de desarrollo.

En términos prácticos, las evaluaciones digitales requieren que cada estudiante trabaje en un dispositivo individual durante un examen. En el modelo actual que se emplea en el PAI se recomienda una conexión a Internet, pero no es obligatoria, ya que los exámenes se cargan previamente en los dispositivos informáticos. El software del examen “bloquea” el dispositivo mientras se ejecuta el examen y registra los momentos en que, por algún motivo, el software sufre alguna interrupción (el supervisor/a puede reiniciar el dispositivo en caso de interrupción justificada del examen). Se recomienda encarecidamente que los exámenes digitales del PAI se realicen en línea, a fin de que se registren las respuestas del alumnado a intervalos regulares y los colegios tengan una menor carga administrativa. Aunque la gran mayoría de colegios y estudiantes utilizan la funcionalidad en línea en el PAI, no es una condición obligatoria para realizar los exámenes, y el IB ofrece un método sin conexión. En el caso de los exámenes digitales del PD y el POP, se requiere una conexión a Internet para realizarlos, y el alumnado accede a ellos a través de un modo restringido que limita el acceso únicamente al examen.

La evaluación digital ofrece ventajas significativas, ya que permite evaluar lo que el IB realmente quiere evaluar, y no simplemente lo que es posible dentro de las limitaciones de los exámenes en papel. Estas ventajas se describen en mayor detalle en la sección “Beneficios de la evaluación digital para la validez” de esta publicación, e incluyen la posibilidad de ofrecer materiales en video, el uso de herramientas de procesamiento de textos para escribir respuestas largas, y la mejora de la accesibilidad, por ejemplo, permitiendo al alumnado ajustar los colores y el tamaño de fuente.

La evaluación digital seguirá evolucionando. A medida que la tecnología avance, el IB irá desarrollando sus evaluaciones para adaptarse a ella. Actualmente, en este tipo de evaluaciones el alumnado solamente utiliza el teclado, el *mouse* (ratón) o el *track pad* (panel táctil) pero, dentro de unos años, es posible que las pantallas táctiles sean lo habitual y abran más posibilidades. Si se mira mucho más adelante, ¿qué papel podría tener la realidad virtual en determinadas evaluaciones? No resulta difícil imaginar que las evaluaciones de lenguas resultarían mucho más auténticas utilizando esa tecnología. Asimismo, las películas de ciencia ficción ofrecen una idea de cómo se podrían gestionar las tareas técnicas (preguntas sobre ciencias) en laboratorios de realidad virtual.

Como parte del compromiso continuo del IB con la transformación digital en sus programas, inevitablemente surgirán nuevas terminologías que afectarán a la comunidad evaluativa en general. Al final de esta publicación se incluye un glosario de términos de evaluación del IB, que incluye términos relacionados con la evaluación digital.

La enseñanza mediante el uso de la tecnología

La tecnología está cambiando la manera en la que se enseña al alumnado: ahora las aulas cuentan con pizarras digitales interactivas, y hay cursos en línea masivos y abiertos que acercan la educación a un público cada vez más amplio. La mayoría del alumnado no escribe sus trabajos a mano, sino utilizando un dispositivo informático.

Precisamente, el paso de la redacción manuscrita al uso de dispositivos provoca una desconexión entre las prácticas de evaluación y las prácticas de enseñanza. Para gran parte del alumnado, el hecho de pasar dos o tres horas escribiendo a mano con un bolígrafo es algo que solo se hace al realizar un examen, por lo que la autenticidad de las evaluaciones se ve seriamente mermada.

Con el aumento del uso de las evaluaciones digitales, es importante que el alumnado esté familiarizado con el estilo de las evaluaciones, para que usar la interfaz no suponga un obstáculo a la hora de demostrar su comprensión de la asignatura. Aunque es posible superar esta dificultad mediante herramientas de familiarización y exámenes de práctica, la solución ideal es que en la enseñanza del alumnado se utilicen las mismas herramientas que en la evaluación.

¿Cómo es una buena evaluación digital?

La evaluación digital puede ofrecer evaluaciones de mejor calidad gracias al uso de la tecnología como medio para plantear preguntas que resultarían imposibles en un examen tradicional en papel. En el IB, una buena evaluación digital:

- Utiliza la tecnología para plantear tareas más válidas, auténticas y pertinentes
- Elimina las barreras administrativas y mejora la realización de los exámenes
- Evita introducir nuevos obstáculos para completar la evaluación

Corrección electrónica

El término *corrección electrónica* se utiliza para denominar cómo el IB corrige trabajos entregados para evaluación mediante el uso de dispositivos informáticos que sirven para mostrar las respuestas del alumnado y registrar las puntuaciones. En el IB, los examinadores/as ya no reciben copias en papel de los exámenes ni de los trabajos de evaluación interna, salvo en circunstancias excepcionales relacionadas con las adecuaciones de acceso e inclusión. El trabajo realizado en papel se escanea para convertirlo a formato electrónico, de manera que los examinadores/as puedan acceder a él por Internet utilizando un software de corrección especializado.

La corrección electrónica supone una serie de grandes ventajas que favorecen la velocidad y la calidad de la corrección:

- El IB siempre dispone del trabajo entregado para evaluación; si por algún motivo un segundo examinador/a tiene que valorarlo, el trabajo está disponible inmediatamente y no es necesario enviarlo por correo postal a otra parte del mundo.
- El IB puede anonimizar el trabajo del alumnado para reducir la posibilidad de que haya sesgos en la corrección que realiza el personal examinador.
- El IB puede implementar un proceso de control de calidad más riguroso, ya que se pueden comprobar los estándares de corrección a medida que esta se realiza.
- Los trabajos entregados para evaluación por parte del alumnado de un colegio pueden repartirse aleatoriamente entre todo el equipo examinador, en vez de que haya un único examinador/a que los corrija todos. De esta manera, se pueden llevar a cabo procesos de control de calidad adicionales.

El IB lleva utilizando la corrección electrónica desde 2010 y este sistema no exige que el alumnado entregue su trabajo digitalmente. Actualmente, la mayoría de los exámenes se contestan a mano y posteriormente se escanean para convertirlos a formato digital.

Riesgos relacionados con la evaluación digital

Desarrollar la evaluación digital requiere ser consciente de varios de desafíos significativos que se deben superar para garantizar un sistema equitativo para todo el alumnado y los colegios. En los siguientes apartados se presentan algunos de los riesgos clave que el IB gestiona. Para disponer de información más detallada sobre este asunto, podrá consultar otras secciones específicas de este documento y otras publicaciones del IB que tratan sobre la evaluación digital.

Carga de trabajo para los colegios

Hay que reconocer que, actualmente, el uso de la tecnología forma parte del mundo del alumnado, y esto debería incluir su experiencia de evaluación. Sin embargo, también está claro que los colegios necesitan tiempo para poder cambiar sus procesos, y que los exámenes digitales requieren unas disposiciones distintas que los exámenes en papel. Por ello, los exámenes digitales del PD y el POP se irán introduciendo de manera que todos los colegios puedan seguir el ritmo del progreso tecnológico, pero sin obligarlos a adaptarse al nuevo sistema de manera demasiado brusca, independientemente de su situación geográfica.

Apoyo a una transición gradual hacia la evaluación digital

El IB trabajará con su comunidad para establecer plazos que permitan que los colegios que lo deseen puedan aprovechar las oportunidades que ofrece la tecnología para realizar una evaluación de mejor calidad. El IB considera que las posibilidades que ofrece la tecnología deben utilizarse de manera inteligente para mejorar la calidad del aprendizaje y la enseñanza y, por consiguiente, de la evaluación.

Fallos tecnológicos

Cuando el alumnado realiza sus exámenes, resulta inadmisibles que haya cualquier tipo de fallo tecnológico. Todo el alumnado debe disfrutar de una experiencia tranquila y sin interrupciones. El IB cuenta con un proceso exhaustivo de familiarización, puesta a prueba y cumplimiento de requisitos que los colegios tienen a su disposición antes del día del examen. La experiencia indica que la mayoría de los problemas que surgen se deben a la infraestructura del colegio. Mediante el proceso de familiarización, puesta a prueba y cumplimiento de requisitos estos problemas pueden identificarse y resolverse con suficiente antelación a los exámenes.

Seguridad

Los exámenes digitales deben ser al menos igual de seguros que los exámenes en papel. Las evaluaciones del IB están diseñadas para que, mientras el alumnado las está realizando, el resto de funcionalidades del dispositivo queden bloqueadas. Cuando el material de evaluación se envía a los colegios antes de los exámenes, como en el caso de las evaluaciones electrónicas del PAI, estos archivos están cifrados y protegidos con contraseña.

Aunque la seguridad debe reforzarse constantemente debido los continuos avances informáticos, no se debe olvidar que el sistema actual de procesos en papel también está expuesto a ciertos riesgos.

Uso razonable de la tecnología

El IB está comprometido con el uso de la tecnología para mejorar la calidad de los elementos que se pueden evaluar; es decir, la tecnología no se utiliza porque sí, sino por las ventajas que aporta. Por lo tanto, el IB seguirá informando sobre cómo añade valor a la validez de sus evaluaciones mediante el uso de la tecnología, y el proceso de desarrollo de la evaluación reflejará este valor añadido.

Sesgos

El IB reconoce que el uso de dispositivos informáticos para realizar las evaluaciones provocará que algunos alumnos/as obtengan mejores resultados que otros. Sin embargo, no hay que olvidar que los exámenes en papel también se ven afectados por algunos sesgos: por ejemplo, el alumnado a quien le resulta difícil escribir con un bolígrafo durante mucho tiempo o que es lento escribiendo a mano está en desventaja.

En todo caso, el IB parte del principio de que no se intenta medir la capacidad del alumnado de utilizar un bolígrafo ni de escribir en un teclado. Por eso, las tareas de evaluación se diseñarán de tal modo que no favorezcan a quienes escriban rápido, ya sea a mano o en teclado.

A fin de poder cumplir este principio, el IB presta mucha atención a las investigaciones que se publican acerca de los efectos del uso de la tecnología.

Cambiar los estándares: ¿arriba o abajo?

Las calificaciones finales del IB tienen un significado; este significado se protegerá mediante el criterio experto del equipo examinador supervisor del IB, y se basará en los datos acerca de los resultados. Por ejemplo, a un alumno/a le puede resultar más sencillo realizar una redacción bien estructurada si puede cortar, pegar y recolocar los párrafos en su escrito a medida que escribe. Esto se tendrá en cuenta al establecer los límites de calificación, para que los alumnos/as que se sientan igual de cómodos con los exámenes en papel y los exámenes digitales obtengan los mismos resultados en lo que resulta importante: reflejar su comprensión y su capacidad analítica en la asignatura.

La evaluación digital permite comprobar nuevos rasgos, así como rasgos que el IB siempre ha valorado pero que anteriormente no se podían evaluar. Por ese motivo, es posible que haya que ajustar las expectativas de lo que se espera que el alumnado demuestre en las evaluaciones (para obtener más información, consulte la sección “Estándares”). No obstante, este cambio se realizará siempre respetando los objetivos arraigados del IB y reflejando las cualidades que se establecen en los descriptores de calificaciones finales vigentes.

Lecturas adicionales

Para obtener más información sobre la evaluación digital en el IB, puede consultar los siguientes recursos:

- La página “Transición a los exámenes digitales” en el Centro de Recursos para los Programas
- *Manual para la evaluación electrónica del PAI*
- “Evaluación electrónica del Programa de los Años Intermedios (PAI) del IB” (video)
- “¡La evaluación electrónica del PAI del IB está lista! Introducción”

La evaluación en la educación

- En el ámbito de la educación, con frecuencia se ha diferenciado entre la evaluación formativa (evaluación **para** el aprendizaje) y la evaluación sumativa (evaluación **del** aprendizaje). En la actualidad, hay una fuerte tendencia en favor de la evaluación **como** aprendizaje.
- El tipo de evaluación que se utilice en un contexto determinado debe depender del uso que se vaya a dar a los resultados, o del propósito de la evaluación.
- La evaluación puede influir en las prácticas de enseñanza, y se debe diseñar para que el efecto de repercusión —es decir, el impacto que la evaluación tiene sobre el aprendizaje y la enseñanza— sea positivo.

Definición de la evaluación

Tal como se explica en la publicación *¿Qué es la educación del IB?* (2019), el Bachillerato Internacional (IB) sigue un enfoque constructivista que le ofrece al alumnado la posibilidad de desempeñar un papel activo y que también reconoce la importancia de contar con un contexto de aprendizaje eficaz (Murphy, 1999). Si la idea es que la evaluación sirva de apoyo para un aprendizaje y una enseñanza eficaces, entonces se debe diseñar basándose en esta teoría constructivista del aprendizaje. Si desea seguir investigando sobre este concepto, puede consultar las obras de Shepard (1992), Wood (1998), Black (1999), y Lambert y Lines (2000).

El término *evaluación* puede referirse a los distintos modos en que se pueden obtener y sopesar datos respecto al desempeño de un alumno/a. Algunos tipos de evaluación frecuentes son los exámenes, los trabajos prácticos extensos, los proyectos, las carpetas, y los trabajos orales. A veces las evaluaciones se llevan a cabo durante un período de tiempo prolongado, mientras que otras veces se realizan durante un período de apenas unas horas. Dependiendo de su tipo y su propósito, los trabajos de evaluación los puede juzgar un profesor/a, otro alumno/a, el propio alumno/a, o un examinador/a.

Figura 5

Ejemplos de distintos tipos de evaluación



Las tareas de evaluación que comprueban el nivel de competencia de un alumno/a pueden utilizar un modelo de **compensación** o un modelo de **dominio** (véase la figura 6).

El modelo de **compensación** se utiliza en la mayoría de exámenes externos y permite demostrar un buen desempeño en un área para mitigar un desempeño más bajo en otra.

Por el contrario, en el modelo de **dominio** de la evaluación es necesario alcanzar un nivel —dominio— mínimo en cada una de las partes de una evaluación.

Un examen consta de tres preguntas, en cada una de las cuales se puede obtener un máximo de 10 puntos.

Si la puntuación necesaria para aprobar fuese un 15, dicha puntuación la podría alcanzar un alumno/a que obtuviese unas puntuaciones de 5, 5 y 5 para cada pregunta, o un alumno/a que obtuviese 10, 3 y 2 puntos respectivamente. En el segundo caso, una puntuación perfecta en la primera pregunta compensa las puntuaciones más bajas logradas en las otras dos preguntas.

En el ejemplo anterior (un examen con tres preguntas, cada una con un máximo de 10 puntos), si para aprobar fuera necesario obtener 5 puntos en cada pregunta, el primer alumno/a aprobaría, pero el segundo alumno/a no. Un tercer alumno/a que obtuviera 10, 10 y 4 puntos tampoco aprobaría.

Figura 6

Comparación entre el modelo de compensación y el modelo de dominio

Modelo de compensación

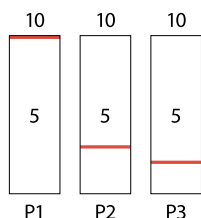
Número de preguntas: 3

Puntuación disponible (en cada pregunta): 10

Puntuación mínima para aprobar (total): 15/30

Si se aplica un modelo de compensación:

La evaluación **se aprueba** debido a que la puntuación total de todas las preguntas es igual o mayor que la puntuación mínima para aprobar.



Modelo de dominio

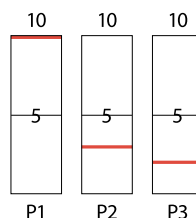
Número de preguntas: 3

Puntuación disponible (en cada pregunta): 10

Puntuación mínima para aprobar (por pregunta): 5/10

Si se aplica un modelo de dominio:

La evaluación **no se aprueba** debido a que la puntuación de las preguntas 2 y 3 no es suficiente.



Las evaluaciones basadas en el modelo de dominio se suelen utilizar en entornos laborales o de formación profesional, donde no resulta adecuado tener muy buenos resultados en un elemento, pero muy malos en otro. Por ejemplo, al hacer una prenda, tener muy buenas habilidades de diseño no puede compensar el hecho de no saber coser.

El IB utiliza varias herramientas de evaluación, entre ellas los exámenes que se deben realizar al final de un programa, y muchas otras tareas en distintas asignaturas que el alumnado completa en distintos momentos y condiciones durante el curso; por ejemplo, ensayos, trabajos de investigación, trabajos escritos, entrevistas orales, investigaciones matemáticas y científicas, proyectos de trabajo de campo, o representaciones artísticas.

Enfoques de la evaluación

La evaluación se puede usar con varios propósitos. El propósito de la evaluación tendrá un gran peso en el diseño de esta. Habitualmente hay dos tipos amplios de evaluación: uno es formativo y el otro es sumativo.

La evaluación formativa es la que está más directamente relacionada con la manera en que el alumnado aprende, por lo cual se la conoce como *evaluación para el aprendizaje*, mientras que a menudo nos

referimos a la evaluación sumativa como *evaluación del aprendizaje*. Esto menoscaba el importante efecto que la evaluación sumativa tiene en lo que realmente se aprende en clase. De hecho, todas las evaluaciones deberían respaldar un aprendizaje adecuado. La evaluación sumativa no consiste únicamente en una actividad que se realiza después de que se haya producido el aprendizaje, sino que se debe diseñar para formar parte integral del aprendizaje y la enseñanza.

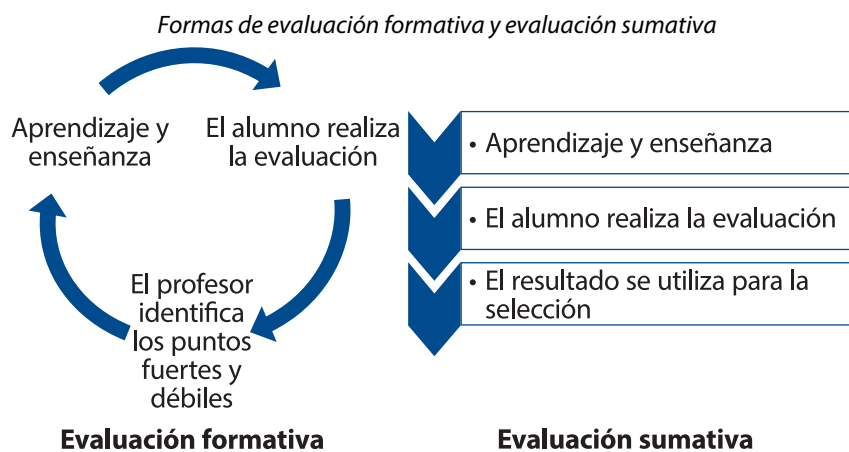
El objetivo general de la evaluación formativa ofrecer información detallada al profesorado y al alumnado sobre los puntos fuertes y débiles del alumnado, para así ayudarle a desarrollar sus capacidades. En este sentido, los tipos de evaluación que implican una interacción directa entre docente y estudiante (como puede ser una discusión) resultan especialmente útiles.

Vygotsky (1962) considera al profesor/a más como alguien que apoya el aprendizaje que como alguien que lo dirige, y para ello debe utilizar tareas e instrumentos de evaluación que ayuden a trabajar al alumnado en lo que denomina *zona de desarrollo próximo*. Esta expresión alude al tramo de logro que abarca desde lo que el alumno/a puede hacer por sí solo hasta lo que puede hacer con el apoyo del profesor/a.

Este concepto es similar a la noción de *andamiaje* de Wood *et al.* (1976). En dicho marco, el profesorado proporciona el andamio para construir el aprendizaje, pero solo cada estudiante puede llevar a cabo dicha construcción. El objetivo del profesorado debe ser establecer evaluaciones formativas que planteen desafíos de un nivel adecuado, y continuar ajustando dicho nivel a medida que el alumnado progresa.

Por su parte, la evaluación sumativa se centra en medir **qué** sabe hacer el alumno/a. Normalmente se utiliza para demostrar que ha completado un programa de capacitación o su grado de preparación para pasar a la siguiente etapa de la educación que recibe. Es decir, la evaluación formativa está interesada en **por qué** el alumno/a hace algo, mientras que en la evaluación sumativa lo importante es saber si el alumno/a ha hecho las cosas **correctamente**. Si bien esto último puede parecer menos útil que saber el porqué, hay que tener en cuenta que el propósito de la evaluación sumativa es emitir un juicio sobre el alumno/a, y no fundamentar la futura enseñanza. En la sección "[Descripción del éxito del alumnado en la evaluación sumativa](#)" de esta publicación se explora en mayor profundidad la necesidad de la evaluación sumativa.

Figura 7



En la evaluación formativa es más importante identificar correctamente el conocimiento, las habilidades y la comprensión que el alumnado aún no ha desarrollado, en vez de medir con precisión el nivel de logro de cada estudiante. El equilibrio entre estas prioridades se conoce como *validez*, y se trata con más detalle en otras secciones de esta guía. En la evaluación sumativa ocurre justo lo contrario a este equilibrio entre el logro del alumno/a y la calidad de los comentarios: los resultados de la evaluación se utilizan para tomar decisiones sobre el alumno/a, con frecuencia relativas a una selección competitiva para oportunidades laborales o educativas, pero también para apoyar la futura enseñanza.

Cualquier análisis de los distintos sistemas de evaluación nacionales pondrá rápidamente de manifiesto la existencia de una gran variedad de técnicas y enfoques de evaluación. Todos estos sistemas tienen sus puntos fuertes y débiles en cuanto a consideraciones técnicas, de recursos y de tiempo, y en cuanto a su influencia en el sistema educativo del país. Incluso si fuese posible, en un contexto dado, empezar

completamente desde cero a diseñar un sistema de evaluación, no existe ninguna técnica que pueda aplicarse por ser considerada la mejor universalmente. Al contrario: las decisiones que se toman al concebir sistemas de evaluación reflejan inevitablemente los valores y las prioridades del contexto social más amplio en el que se adoptan. Para profundizar en este tema, consulte Cresswell (1996) y Broadfoot (1996).

También es importante ser consciente de que la evaluación sumativa se utiliza cada vez más como medida de la calidad de la enseñanza, lo cual añade otra dimensión a por qué se realiza la evaluación: para el beneficio del sistema educativo más que el del alumno/a.

El efecto de repercusión y el aprendizaje

Lo que se necesita es un proceso de evaluación que sea lo más válido posible [...]. Sin embargo, dicho proceso no debe, por su "efecto de repercusión", distorsionar la buena enseñanza ni ser demasiado lento, ni absorber demasiados de nuestros escasos recursos educativos.

(Peterson, 1971)

En esta cita, Alec Peterson, fundador del IB, reconoce el riesgo de que el modo en que se realiza la evaluación pueda influir en el enfoque del aprendizaje. Como Surgenor (2010) indica, esto puede conllevar implicaciones tanto negativas como positivas para la experiencia de aprendizaje y enseñanza del alumnado.

En una obra de 1971, Snyder propuso que el alumnado creara su propia comprensión del currículo basándose en mensajes implícitos y explícitos sobre los elementos que cuentan en la evaluación, lo cual se denomina *currículo oculto*. Esto puede ayudar al alumnado a entender cómo aprobar una asignatura, pero no sirve para comprender la asignatura. Gibbs presenta otra forma de entender esto en la siguiente cita:

Cuando volví a realizar el examen, me centré solamente en aprobarlo. Obtuve una puntuación del 96 %, y aquel tipo no entendía por qué yo no había aprobado la primera vez. Le dije que en esa segunda ocasión solamente me concentré en aprobar el examen, en vez de en comprender la asignatura. Aún sigo sin comprender la asignatura, así que se puede decir que el examen no logró su objetivo.

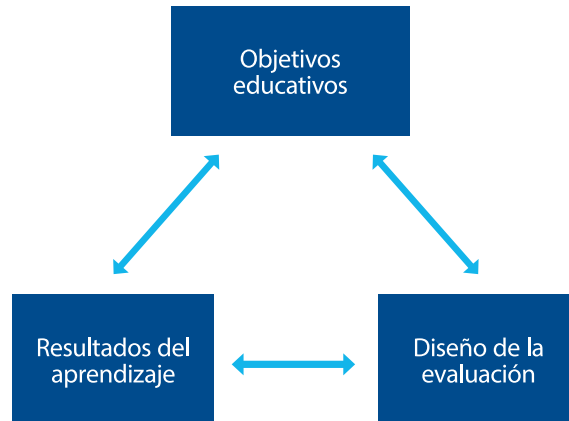
(Gibbs, 1992, p. 101)

El concepto de *repercusión* también se plasma en la frase tan oída de "Si no se va a evaluar, no se enseña". Esto destaca que la evaluación y la enseñanza no se pueden considerar como conceptos independientes entre sí.

La evaluación debería servir para aportar pruebas de que el alumnado ha alcanzado los objetivos educativos, pero hay varios motivos por los que es posible que no lo logre. Lo más probable es que los resultados del aprendizaje en los que se basa una evaluación no reflejen de manera eficaz el propósito inicial de la educación. Un ejemplo sería un curso de formación profesional para entrenar a atletas de élite cuyos resultados del aprendizaje consistieran en comprender las reglas del atletismo. Otro problema común es que, mientras la evaluación puede cubrir todos los resultados propuestos, la mayoría de las pruebas de examen pueden centrarse en tareas que resulta fácil evaluar, pero que no son particularmente importantes para los objetivos educativos generales.

Esta interdependencia de la evaluación y el propósito educativo se puede observar en la figura 8, creada a partir del paradigma de Furst (1958), citado en Frith y Macintosh, (1984). Si se produce una desconexión entre dos de estos tres elementos, casi con total seguridad la evaluación tendrá poca calidad.

Figura 8
Relación entre objetivos, resultados y diseño



Validez

- El término *validez* se refiere a que la evaluación sea **adecuada al propósito**. La validez es un concepto complejo que tiene muchos aspectos.
- Las evaluaciones se usan con diversos propósitos.
- Una evaluación puede resultar válida para un propósito, pero no para otro. Por ejemplo, una prueba de ortografía no sirve también para medir la fluidez en un idioma.
- Se podría argumentar que lo que sería válido, o no, sería el propósito relacionado con los resultados de la evaluación, en lugar de la evaluación en sí.
- La principal preocupación del IB es si el programa es válido; en segundo lugar, si los elementos del programa (como los cursos individuales) son válidos y, finalmente, si la evaluación individual es válida.

Definición de la validez

¿Cuál es el propósito de realizar exámenes? ¿Para qué sirven? Estas preguntas conforman el núcleo de lo que significa que una evaluación sea “adecuada al propósito” o que “mida lo que debe”. Además, frecuentemente las distintas respuestas a estas preguntas están en conflicto entre sí. Puede darse también el caso de que no todos los propósitos de una educación puedan evaluarse en un examen en concreto, o que no sea posible evaluar algunos propósitos.

A modo de ejemplo, pongamos que se ofrecen las cuatro posibilidades siguientes como propósito de una evaluación en matemáticas:

1. Determinar qué comprende el alumno/a una vez concluido el programa de estudios
2. Ser un medio de selección para estudios posteriores o para el mundo laboral
3. Ser un indicador para saber si el alumno/a tendrá éxito en el futuro
4. Reforzar la enseñanza de los objetivos curriculares del programa

Resulta difícil conciliar incluso estos objetivos sencillos: si el éxito futuro en las matemáticas depende más del análisis que de la geometría, ¿debería la evaluación centrarse en el análisis? ¿Qué relación guarda esto con el primer objetivo de determinar qué sabe el alumno/a (en geometría)? ¿Qué ocurre cuando dos o más de estos objetivos dan lugar a un conflicto sobre cómo debe diseñarse una evaluación?

Estos cuatro posibles propósitos no son definitivos ni exhaustivos. Para Newton (2007), los resultados de la evaluación pueden tener diversos usos, y enumera una serie de ejemplos:

<ul style="list-style-type: none"> • Evaluación social • Formativos • Supervisión del alumnado • Transferencia • Determinación del curso adecuado para un alumno/a • Diagnóstico • Orientación • Medición de aptitudes • Selección 	<ul style="list-style-type: none"> • Obtención o renovación de licencias • Elección de colegio • Supervisión de la institución • Asignación de recursos • Intervención organizativa • Evaluación del programa • Supervisión del sistema • Equiparabilidad
---	---

A pesar de esta diversidad de usos que entran en conflicto entre sí, sigue siendo necesario que individuos e instituciones tomen decisiones sobre las personas en diversos contextos. Si no tienen disponibles los resultados de la evaluación, se corre el riesgo de que utilicen otras maneras peor diseñadas y más sesgadas para tomar dichas decisiones. Cresswell (1986) planteó el problema de la siguiente manera:

Claramente, si los otros criterios son menos fiables que los exámenes, confiar más en dichos criterios provocará que se tomen decisiones de selección menos fiables.

(Cresswell, 1986, p. 42)

La noción de *validez* engloba el concepto de que un título académico (o parte de este, como podría ser una evaluación individual) sea adecuado al propósito para el que está pensado.

El término **validez** se usa a menudo como elemento esencial de la calidad de una evaluación, pero hace mucho tiempo que los académicos/as intentan darle una definición precisa (Newton, 2012). Además, hay una complicación añadida: hay un concepto más restringido de la validez, que se utiliza a menudo con el término **fiablez**, para referirse a la capacidad de una evaluación para medir realmente lo que debe medir. Para evitar confusiones, esta publicación se refiere a este segundo concepto como *pertinencia del constructo*. Véase la sección "[Pertinencia del constructo y autenticidad](#)".

En última instancia, determinar la validez de una evaluación es una cuestión de juicio. Este juicio se basa en la obtención sistemática de pruebas que respalden las interpretaciones de las puntuaciones obtenidas en los exámenes para los usos propuestos de dichos exámenes. Las demandas opuestas de una evaluación a las que nos hemos referido anteriormente en esta sección se explicarán con más detalle más adelante, pero siempre habrá que llegar a un punto de equilibrio entre dichas demandas. La cuestión importante es si las pruebas disponibles indican que la evaluación es lo suficientemente adecuada al propósito como para resultar útil. En consecuencia, al hablar de validez es recomendable reflexionar sobre la solidez de las pruebas aportadas —o del argumento de validez que se plantea—, en lugar de buscar una respuesta sencilla acerca de si la evaluación es válida o no. Por tanto, decidir el grado de validez de algo es un juicio basado en pruebas. La validez es contextual y está orientada a un propósito, y no un atributo binario. Newton y Shaw explican que se debe considerar en qué medida una evaluación es adecuada para su propósito antes de realizar una afirmación sobre su validez, y que "la validación se considera el proceso de desarrollar y sopesar la solidez de un argumento relativo a la interpretación y al uso de las puntuaciones obtenidas en los exámenes" (Newton y Shaw, 2014, p. 3). Por lo tanto, la validez no es una propiedad de la evaluación en sí, sino de las inferencias que se extraen de los resultados de la evaluación.

Por último, la validez no es algo que se logre de manera aislada durante el diseño de una evaluación, sino que se desarrolla continuamente durante su ciclo de vida. Asimismo, el argumento de validez no se formula al principio del ciclo de vida, sino que se amplía continuamente y se va refinando durante el proceso de diseño de la evaluación.

Creación de un argumento de validez

La validez no es un concepto objetivo sencillo, sino que se trata de un punto de equilibrio entre cuestiones en conflicto. No se puede "demostrar" la validez, sino que se debe preparar un argumento convincente sobre por qué las decisiones tomadas sobre un curso o una evaluación los hacen significativos y adecuados al propósito que se persigue. En consecuencia, lo válido no es la evaluación (o el resultado), sino el propósito al que se somete. Newton (2012) reflexiona que "el uso de un procedimiento de evaluación concreto para un propósito específico (es decir, tomar una decisión específica) resulta válido si su argumento interpretativo es suficientemente sólido".

Por razones similares, las pruebas de los argumentos de validez deberían surgir de manera natural de los procesos de desarrollar, supervisar e impartir un programa. Además, deberían reflejar las discusiones y las decisiones alcanzadas durante dichos procesos. Un ejemplo en la evaluación sería que los juicios y las valoraciones que hay tras la decisión de qué preguntas se harán en un examen en concreto demostrarán que el currículo se ha cubierto de manera adecuada.

Es esencial que el argumento de validez tenga una estructura adecuada que cubra todos los aspectos que se consideran importantes. En el IB, esto está representado por una serie de preguntas para las que se recopilan pruebas.

Mantenimiento de la validez

La validez no es algo que se pueda decidir o “demostrar” al diseñar un modelo de evaluación, sino que es algo que se va desarrollando con la evaluación durante su ciclo de vida y mientras que las decisiones se tomen basándose en el resultado de la evaluación.

Durante la vida de un curso concreto, el IB recabará distintas pruebas que demuestren que una evaluación determinada es válida (es decir, distintos argumentos de validez).

Una vez desarrollado o revisado el curso, las discusiones sobre el propósito del curso y sobre cómo se debería evaluar para cumplir con su propósito conformarán el núcleo del argumento de validez. Se tendrán especialmente en cuenta el equilibrio entre la pertinencia del constructo y los demás aspectos de la validez. En cada convocatoria de exámenes, la redacción de un nuevo examen o tareas de evaluación generará más pruebas, especialmente en lo que respecta a la fiabilidad, la equidad, la equiparabilidad, la manejabilidad, y la pertinencia del constructo. De este modo, los comentarios del profesorado y los resultados del alumnado servirán para determinar en qué medida la evaluación alcanza sus objetivos, así como para proporcionar pruebas de la fiabilidad de la corrección y de la equiparabilidad de las calificaciones. Esta información se utilizará para desarrollar el siguiente grupo de evaluaciones.

Finalmente, las pruebas recabadas de todas las evaluaciones de un curso concreto, y de todos los cursos de un programa específico, sirven de base para las decisiones de la siguiente revisión del curso o del programa.

Evaluar la validez de un examen no es algo estático que solo sucede una vez, sino que es un proceso continuo.

(Sireci, 2007, p. 477)

Elementos de la cadena de validez

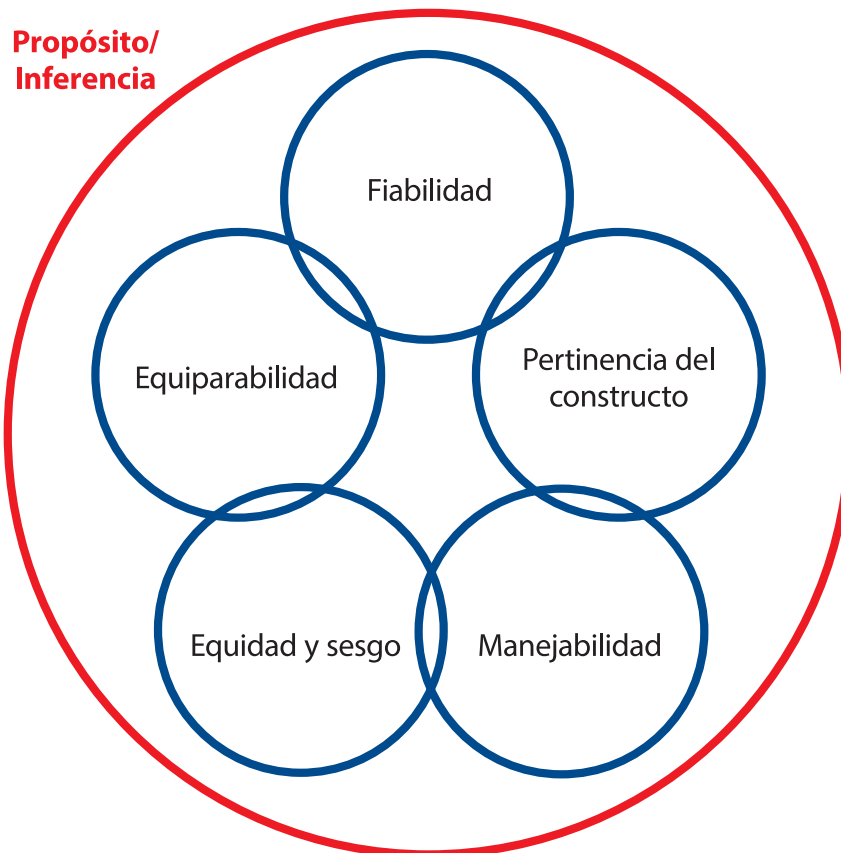
- La validez implica elementos muy diversos. Si se piensa en ellos como si fueran una cadena, se puede decir que si un eslabón está roto, también lo estará toda la evaluación.
- Normalmente, el IB se centra en cinco elementos de la validez: fiabilidad, pertinencia del constructo y autenticidad, manejabilidad, equidad y sesgo, y equiparabilidad.
- Con frecuencia, estos cinco elementos están en tensión los unos con los otros. Al considerar si una evaluación es válida, es necesario tener en cuenta el propósito principal de la evaluación para determinar la importancia relativa de cada elemento.
- El IB pone mucho esmero en crear programas, cursos y evaluaciones que demuestren pertinencia del constructo.
- En el ámbito de la evaluación, esto significa que el IB hace hincapié en ofrecer tareas y preguntas significativas que midan las habilidades de pensamiento de orden superior, en vez de en crear exámenes que se puedan corregir de manera fiable con facilidad.

La validez y la fiabilidad se consideran características esenciales de cualquier sistema de evaluación, especialmente si es de alto impacto, en el que el resultado es muy importante para el alumnado o el profesorado. Estas características son polifacéticas, ya que hay distintos tipos de validez y fiabilidad.

La idea de la **cadena** muestra la complejidad de la validez (Crooks *et al.*, 1996). Cada uno de los cinco elementos de la cadena son importantes en sí mismos, pero por sí solos no son suficientes para hacer que una evaluación sea válida (es decir, adecuada al propósito). Por ejemplo, es posible que una evaluación sea muy fiable, pero que sistemáticamente ponga en situación de desventaja a un grupo concreto. Asimismo, es posible que una tarea se centre exactamente en aquello para lo que estaba pensada y que, debido a su duración y requisitos, también evalúe la resistencia del alumnado y el ingenio del colegio. Las siguientes secciones cubren con más detalle cada uno de estos elementos. Cabe recordar que, si bien todos los elementos son necesarios para conseguir la validez, entre ellos hay tensiones. Por ejemplo, si se quiere que un examen cubra todos los aspectos del currículo, esto puede provocar que las pruebas sean demasiado largas para el alumnado. Del mismo modo, contextualizar una prueba para cada cultura o país concreto en el que se realiza (equidad) puede plantear dudas sobre si realmente se aplica la misma prueba en cada caso (equiparabilidad).

Para que una evaluación sea válida, la cadena debe constar de todos sus eslabones.

Figura 9
La cadena de validez






El equilibrio entre los distintos aspectos de la validez

Al diseñar una evaluación, es importante hallar un equilibrio entre los distintos aspectos de la validez, normalmente en el contexto del propósito de la evaluación. Una misma evaluación no puede conseguir los máximos estándares en cada uno de estos elementos, así que hay que llegar a un punto medio de equilibrio. Del mismo modo, algunos de estos elementos se fijan durante el diseño de la evaluación, pero otros (particularmente la fiabilidad y la equidad) evolucionan durante el tiempo que se realiza la evaluación y durante el período de corrección y calificación.

Figura 10

Hallar un equilibrio entre prioridades contrapuestas de dos aspectos de la validez: equidad y manejabilidad.

		
Falta de equilibrio: Carga de trabajo poco razonable sobre el alumnado (manejabilidad)	Equilibrio correcto: Se evalúa al alumnado de manera razonable	Falta de equilibrio: La suerte tiene un papel demasiado destacado en la decisión de qué tema se evalúa (equidad)
La evaluación contiene 50 tareas extensas (cada una de 20 minutos de duración) que cubren todos los aspectos del curso. Esto significa que la evaluación dura más de 16 horas.	La evaluación contiene 20 preguntas cortas (2 minutos cada una), 5 tareas más largas (10 minutos cada una) y 2 tareas en profundidad (30 minutos cada una). Cada pregunta trata sobre un aspecto distinto del curso. La evaluación en total dura dos horas y media y cubre algo más de la mitad de los posibles temas del curso.	La evaluación contiene una tarea en profundidad de 30 minutos en la que se miden tanto el conocimiento como la comprensión de un aspecto del curso (un aspecto de entre 50 posibles). La calificación final se basa íntegramente en una tarea.

A pesar de que tienen definiciones diferentes, hay considerables coincidencias en la manera en que estos cinco elementos se dan y se gestionan. En conjunto, forman el concepto más amplio de *validez* en las evaluaciones del IB.

Por último, el IB da la mayor prioridad a la **validez de constructo**, es decir, que sus evaluaciones midan los rasgos y las competencias que se quieren medir. Sin embargo, es importante recordar que esta priorización no puede producirse completamente a expensas de los otros aspectos de la cadena de validez.

Fiabilidad

La fiabilidad en la evaluación puede definirse como la medida en que un alumno/a obtendría el mismo resultado en una prueba si se repitiera el procedimiento empleado. Esto no es necesariamente lo mismo que decir que el alumno/a obtendría el resultado “correcto”.

En su introducción al concepto de fiabilidad, Winkley y Cresswell (2011) enumeran las áreas en las que la fiabilidad es deseable —como al valorar el desempeño del alumnado—, junto con ejemplos que destacan una lista de posibles amenazas a la fiabilidad.

1. **Fiabilidad entre las distintas correcciones.** Un examinador/a puede ser más o menos tolerante que otro en determinadas preguntas. De manera similar, puede que el mismo examinador/a se muestre más o menos tolerante de un día para otro.
2. **Variabilidad en el desempeño del alumnado.** Es posible que el desempeño de un alumno/a en un examen varíe de un día a otro, sobre todo si cambian las condiciones en las que se realiza el examen. Algunos ejemplos de dichas condiciones podrían ser si el examen es por la mañana o por la tarde, quién es la persona que administra el examen, si el alumno/a ha dormido bien la noche anterior, o si desde el aula se oye a alguien cortando el césped.
3. **Distintos exámenes.** Las preguntas serán diferentes entre un examen y otro, y puede que evalúen diferentes facetas de la comprensión del alumnado (normalmente los exámenes tratan sobre algunos temas del currículo, ya que no hay suficiente tiempo para evaluarlos todos, y es posible que el alumnado decida estudiar algunos temas concretos y no otros).
4. **Equiparabilidad de los resultados de un año a otro.** Esta equiparabilidad garantiza que los resultados del alumnado puedan compararse de forma significativa entre años.

5. **Diferencias entre las especificaciones de los exámenes.** Garantizar la equiparabilidad a lo largo del tiempo puede resultar complicado cuando se producen cambios en las especificaciones de los exámenes o en los programas de estudios.
6. **Distintos tipos de actividades de evaluación.** Muchas evaluaciones están compuestas por diversos tipos de actividades. Esos métodos variados de evaluación pueden implicar diversos problemas de fiabilidad.
7. **Distintos tipos de preguntas.** Es posible que el alumnado obtenga diferentes resultados en función del tipo de preguntas que se le presenten.

En la práctica, el IB normalmente prevé que un alumno/a solo realiza la evaluación una vez y busca la **coherencia** en los resultados del proceso. Por tanto, el IB se centra en los puntos 1, 3, 6 y 7 de esta lista.

El IB es una organización educativa y, como tal, toma medidas para aumentar los niveles de fiabilidad en sus evaluaciones. En este sentido, la fiabilidad en la corrección es un aspecto central; de hecho, muchos de los propósitos de la estandarización, el modelo de calidad y la moderación consisten en asegurarse de que el personal examinador realice la corrección de manera coherente.

Fiabilidad de la corrección

El siguiente ejercicio rápido demuestra el concepto de fiabilidad de la corrección y cómo puede funcionar en la práctica.

En la figura 11 se muestran cinco fragmentos de respuestas reales de una evaluación de Adquisición de Lenguas del IB.

Use el esquema de calificación que se muestra a continuación para evaluar sobre 8 puntos la calidad del lenguaje, y registre las puntuaciones que otorgaría a cada trabajo.

Una vez que haya corregido cada trabajo, vea los comentarios sobre la calidad de su corrección. Los comentarios aparecen después de los exámenes con las respuestas del alumnado.

Esquema de calificación

¿Con qué eficacia y precisión aplica el alumno/a las habilidades de reflexión?

Puntuación	Descriptor de nivel
0	El alumno/a no alcanza ninguno de los niveles especificados por los descriptores que figuran a continuación.
1-2	El alumno/a indica la repercusión del proyecto en su aprendizaje y si se han logrado los objetivos generales.
3-4	El alumno/a resume la repercusión del proyecto en su aprendizaje y si se han logrado los objetivos generales, y lo respalda parcialmente con ejemplos.
5-6	El alumno/a describe la repercusión del proyecto en su aprendizaje y evalúa el resultado según los criterios de logro, y lo respalda con ejemplos.
7-8	El alumno/a explica la repercusión del proyecto en su aprendizaje y evalúa el resultado según los criterios de logro, y lo respalda completamente con pruebas específicas y/o ejemplos detallados.

Vea cada una de las respuestas del alumno/a y asígneles una puntuación usando el esquema de calificación como guía.

Figura 11A

Ejercicio sobre la fiabilidad de la corrección

Como seres humanos, estamos libres de hacer lo que sea; cualquier cosa. Se entiende que esto puede ser ~~un~~ visto de varias perspectivas. Personalmente, yo valoro el respeto como lo más importante en el momento de decidir hacer algo. Respeto la opinión, la cultura y la religión de otra persona antes de hacer algo, porque sé que lo que yo elijo hacer podría tener consecuencias. Entonces, creo que todos ustedes saben la importancia de tener en cuenta los sentimientos de tus compañeros antes de decirte algo, ¿verdad?

Figura 11B

Ejercicio sobre la fiabilidad de la corrección

Por un lado somos libre para hacer lo que quiera, no hay restricciones como en otros país, ~~pero~~ podemos ir al compras comer y dormir lo que ~~quiero~~ quiera pero no soy libre para hacer los demás sufrir, por ~~ej~~ ejemplo no soy libre a causa alguien a sufrir muerte. Por otro lado soy libre decir lo que quiera pero si causa alguien a sufrir a causa de nos palabras por supuesto hay consuencias por nos acción en eso caso, la persona dice palabras malas.

Figura 11C

Ejercicio sobre la fiabilidad de la corrección

En primer lugar estoy de acuerdo con el hecho que las acciones y las decisiones que alguien toma no debería ser validas si hace sufrir a los demás. Pero infelizmente vivimos a una sociedad donde palabras pueden offendir, acciones hace sufrir y la mayoría del tiempo las personas son bastante sensibles. Hay veces donde lo que el mundo necesita es mas amor y menos accoso.

Figura 11D

Ejercicio sobre la fiabilidad de la corrección

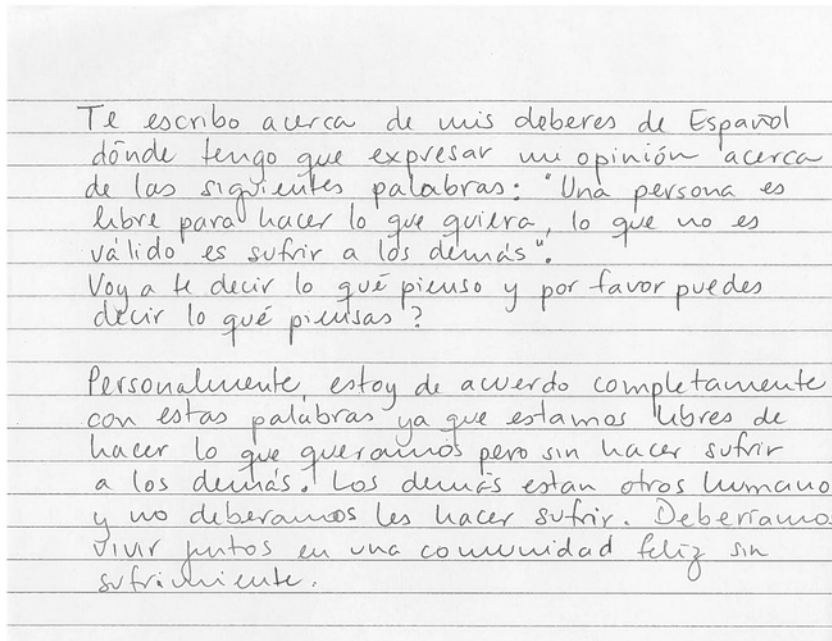
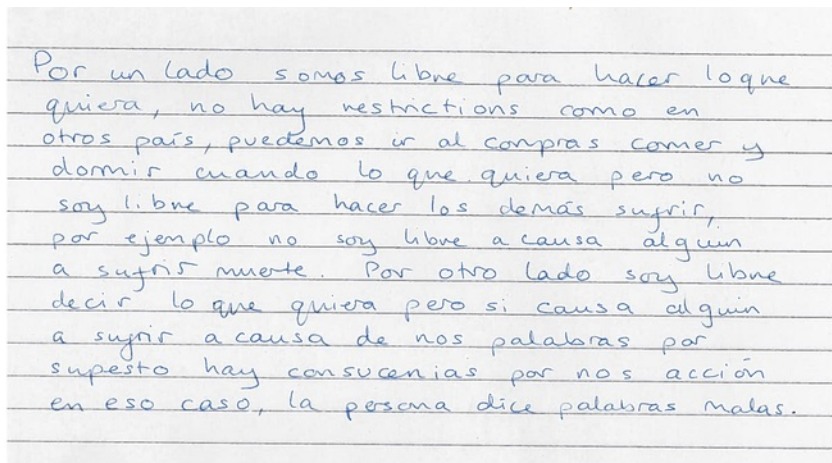


Figura 11E

Ejercicio sobre la fiabilidad de la corrección



Una vez que haya finalizado la corrección de cada examen y haya otorgado una puntuación a cada alumno/a, vea los comentarios sobre su corrección.

Comentarios

¿Ha observado que los trabajos de los alumnos/as B y E son idénticos, pero están escritos con una caligrafía diferente? ¿Ha otorgado la misma puntuación al alumno/a B que al alumno/a E?

Quizás también se haya encontrado ante la duda de cuál de las dos puntuaciones de una banda de puntuación en particular debería otorgarse. En el caso de que tuviera que corregirlos de nuevo, ¿con qué seguridad tomaría la misma decisión?

Este ejercicio no pretende hacer hincapié en la calidad del trabajo del alumnado, sino poner de manifiesto que una misma persona puede hacer valoraciones un tanto diferentes cada vez que corrige un trabajo. En

líneas generales, su opinión sobre la calidad del trabajo es la misma cada vez; sin embargo, la puntuación exacta puede variar levemente.

Repita el ejercicio corrigiendo cada trabajo de nuevo pasados unos días, a fin de observar las similitudes entre las puntuaciones. (Trate de no acordarse de las primeras puntuaciones que otorgó).

Para obtener más información sobre cómo el IB asegura la fiabilidad de la corrección, consulte el video [“Fiabilidad: cómo garantizamos la ecuanimidad en la corrección”](#).

Resultados coherentes y resultados “adecuados”

El objetivo de lograr un nivel alto de fiabilidad es que el alumnado obtenga la misma puntuación (justa) sea quien sea el examinador/a que corrija sus trabajos, y no una puntuación “adecuada”. Para determinar la calidad de un trabajo, el profesorado emplea su juicio profesional, pero con frecuencia cuesta que dos docentes estén de acuerdo sobre qué puntuación conceder. Se considera que hay fiabilidad cuando ambos/as se rigen por el mismo juicio o criterio (el del examinador/a supervisor).

Esto supone un desafío particular al abordar las consultas sobre los resultados. En estos casos, el examinador/a se tiene que asegurar de que la puntuación que conceda sea la misma que hubiera otorgado al primer examen que corrigiera, y debe evitar dejarse influir (positivamente o negativamente) por información adicional que ahora pudiera tener, como los límites de calificación o las consecuencias para el alumno/a.

Se ha documentado ampliamente que fuera del sector de la educación se entiende muy poco el concepto de fiabilidad de la evaluación, pero este es un tema en el que hay que hacer más hincapié, ya que cada vez hay más debate público sobre los resultados de la evaluación.

Tal como la bibliografía apuntó que pasaría, los participantes consideraron que la fiabilidad de la evaluación, y en particular la imprecisión de las mediciones, eran conceptos difíciles de comprender.

(Chamberlain, 2010, p. 3)

A veces pienso que los exámenes no son un sistema justo para medir la capacidad de una persona, ¿verdad? En realidad, reflejan más el estado de esa persona en ese momento, y lo estudioso que era y esas cosas. Alguien a quien no se le den muy bien los exámenes podría destacar más adelante y convertirse en el primero de su área. (Empleado del sector de la salud, varón).

(Chamberlain, 2010, p. 27)

Aproximadamente el 63 % del profesorado y alumnado eligieron la opción “Cualquier nivel de error debe ser inaceptable, incluso que un único alumno reciba una calificación incorrecta es completamente inaceptable” por un lado, mientras que más de la mitad de ellos también eligieron “Hay una diferencia entre un error evitable, como una falta de ortografía en un examen, y algo inevitable, como la falta de coherencia entre dos examinadores”, lo cual sugiere una tolerancia hacia los errores. Esta falta de coherencia refleja la débil relación entre el conocimiento de la fiabilidad y las actitudes hacia la falta de fiabilidad.

(He *et al.*, 2010, p. 27)

Pertinencia del constructo y autenticidad

¿Con qué precisión se mide aquello que se intenta medir? La medición precisa de las habilidades y los conocimientos que se quieren comprobar se denomina a veces *validez de constructo*. Sin embargo, para evitar que haya confusiones con el término general de *validez*, en esta publicación se utiliza la expresión *pertinencia del constructo*.

La idea de una evaluación auténtica está estrechamente relacionada con la pertinencia del constructo. **Autenticidad** significa que, siempre que sea práctico, la prueba se realiza de manera que se ajuste a las situaciones en las que el alumnado podría esperar encontrar el problema en el mundo real. Las evaluaciones que aíslan las tareas del contexto en el que aparecen, que simplifican en exceso o que resultan forzadas son ejemplos de evaluaciones que no son auténticas.

Es sencillo encontrar casos de malos diseños de evaluación que tienen poca pertinencia del constructo; por ejemplo, utilizar un examen oral para medir la capacidad de escribir una carta. Sin embargo, a menudo

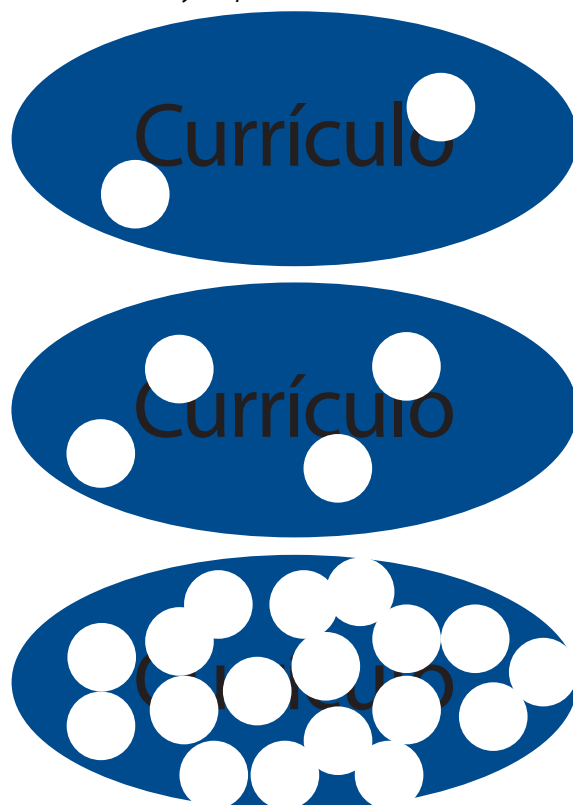
incluso un enfoque ampliamente aceptado para establecer las tareas no permite evaluar realmente lo que se pretende. A modo de ejemplo, pensemos de qué manera se preparan los exámenes de literatura tradicionales. Al alumnado se le han enseñado los recursos literarios que se utilizan en un texto concreto y, después, la evaluación le pide que redacte un ensayo relacionado con uno de esos recursos. Si el alumnado se limita a recordar y reproducir todo aquello que les explicó el profesor/a, ¿realmente está demostrando su comprensión de la literatura?

Comprender claramente qué se quiere que una tarea de evaluación permita demostrar al alumnado, y analizar meticulosamente en qué medida le ha permitido demostrarlo, representan las mejores prácticas a fin de garantizar unos niveles adecuados de pertinencia del constructo. Plantearse qué otras habilidades necesitará el alumnado para poder llevar a cabo la tarea es especialmente útil. Las tareas abiertas, como los ensayos o proyectos, resultan especialmente proclives a exigir que el alumnado tenga una capacidad muy buena de redacción para demostrar las habilidades de investigación o de análisis que la tarea pretende abordar. Esto queda englobado en el concepto de *diseño universal de evaluación*, que pretende reducir las barreras al alumnado en los casos en que las dificultades no estén relacionadas con lo que la evaluación debe evaluar (Dolan *et al.*, 2013).

La pertinencia del constructo está estrechamente vinculada con el **currículo**. Muchas veces, una evaluación solamente planteará preguntas basadas en una selección reducida de materiales curriculares. La cuestión de si el currículo se ha cubierto de manera suficiente es un tema de pertinencia del constructo y depende del juicio profesional. Del mismo modo, elegir los tipos de preguntas que se van a plantear es una parte importante de este diseño de evaluación, ya que los distintos tipos de preguntas o tareas pueden evaluar distintos tipos de constructo.

Figura 12

Los exámenes y en qué medida se cubre el currículo



Manejabilidad

El elemento de la manejabilidad no se ha investigado a fondo, ni tampoco existe una única definición o enfoque para medir este parámetro. En términos generales, la *manejabilidad* se refiere al esfuerzo necesario para realizar la evaluación desde el punto de vista del alumnado, del colegio y del IB.

Para el **alumnado**, la manejabilidad suele estar relacionada con el esfuerzo necesario para completar una evaluación. Por ejemplo, se considera que un examen de 8 horas de duración no resulta razonable para una persona de 18 años. De manera similar, al valorar la manejabilidad, también hay que tener en cuenta el momento en que se realiza la evaluación y la duración del período de exámenes en un día determinado para un alumno/a en concreto. El IB debe tener en cuenta la carga de trabajo acumulada que suponen otras evaluaciones del IB que es posible que el alumnado realice como parte de su educación. Esto está relacionado con el principio de que el IB considera la validez de los **programas**, y no solo de las **asignaturas** individuales.

Es posible que a un **colegio** se le requiera proporcionar material para una evaluación, lo cual afecta a la manejabilidad de dicha evaluación. Por ejemplo, en un curso de formación profesional de ingeniería puede que sea necesario que cada estudiante disponga de un motor que pueda montar. Dependiendo de la cantidad de estudiantes, la cantidad de motores que se necesitan puede resultar imposible de manejar.

Otro aspecto de la manejabilidad para los colegios tiene que ver con cómo se hace llegar al IB el trabajo del alumnado para su corrección. La obligación de grabar en audio o video una presentación implica para el colegio una mayor exigencia que entregar un trabajo escrito.

Por último, el **IB** también tiene en cuenta la manejabilidad en lo que se refiere a la cantidad de trabajos entregados para evaluación que se revisan. Por ejemplo, si bien asistir a una producción teatral de tres horas puede aportar las mejores pruebas de la capacidad de un alumno/a, esta no es una opción práctica si la tarea se tiene que revisar externamente. Ofrecer una amplia variedad de preguntas opcionales también podría afectar a la manejabilidad, ya que establecer un estándar común entre ellas requiere mucho trabajo y mucho tiempo.

Con frecuencia hay cierta tensión entre la **manejabilidad**, la **fiabilidad** y la **pertinencia del constructo**. Por ejemplo, aumentar la cantidad de evaluaciones realizadas —o ampliar la duración de un examen determinado— permitiría al alumnado aportar más pruebas de su comprensión del currículo completo. Esto aumentaría la probabilidad de que las decisiones de corrección demasiado generosas y las demasiado severas se acaben anulando las unas a las otras y queden compensadas, lo cual mejoraría la fiabilidad. Sin embargo, esto también provocaría que la evaluación midiese la capacidad del alumnado para mantener su desempeño en un examen largo, en vez de los objetivos de la evaluación, lo cual disminuiría la pertinencia del constructo y la manejabilidad.

El IB realiza controles estrictos a la manejabilidad de la evaluación, en particular en lo que se refiere a la cantidad total de evaluaciones en cada curso.

Equidad y sesgo

Se considera que una prueba está sesgada cuando da una ventaja o una desventaja involuntaria a un alumno/a. A continuación se plantean algunos ejemplos de exámenes con sesgos:

- Un examen de Historia escrito completamente en latín
- Un examen de Matemáticas en el que todas las preguntas tratan sobre apuntar carreras en un partido de críquet
- Un examen de Arte en el que el caballete para realizar una pintura coloca a dos metros del suelo

En todos estos casos, las tareas pueden poner a parte del alumnado (tal como a los alumnos/as más bajos en el último ejemplo) en una situación de clara **desventaja**. En la práctica, la mayoría de ejemplos de sesgos son más sutiles que los casos anteriores, pero, a no ser que se preste atención, los sesgos pueden marcar grandes diferencias en los resultados del alumnado. Contextualizar las preguntas sin generar ningún sesgo es especialmente complicado, ya que algunas situaciones pueden resultar familiares a una parte del alumnado, pero no a otra, sobre todo teniendo en cuenta la naturaleza internacional del IB.

El *sesgo* puede definirse como la diferencia que se produce en el resultado de un proceso de evaluación y que no tiene que ver con una auténtica diferencia de la aptitud o de los logros que se miden. Los sesgos pueden surgir de las condiciones en que se **realiza** la evaluación, de la **corrección** de la evaluación (lo que se convierte en un problema de fiabilidad de la corrección), o de las propias **tareas** de evaluación.

Sesgos relacionados con las condiciones en que se realiza la evaluación

Los ejemplos que aparecen al principio de esta sección incluyen un sesgo relacionado con las condiciones en que se realiza la evaluación: la improbable circunstancia de que el caballete estaba demasiado alto para parte del alumnado. Sin embargo, hay muchas maneras en que las condiciones en que se realizan las evaluaciones, y en concreto los exámenes, pueden implicar sesgos. El sesgo más común tiene que ver con el horario en el que se realiza un examen. Los exámenes que se realizan durante períodos de temperaturas o de contaminación extremas en algunas partes del mundo, o durante períodos en los que el alumnado no puede concentrarse o prepararse de manera adecuada, implicarían sesgos. Esto supone un desafío particular para el IB, ya que los requisitos contrapuestos de los distintos países y regiones (territorios) hacen que cualquier fecha posible suponga problemas para al menos algunos colegios.

La distribución de la sala donde se realiza el examen puede constituir también un sesgo. Pongamos el ejemplo de dos alumnos/as que realizan el examen, uno sentado donde le da la luz del sol directamente, y el otro en una zona con sombra. La falta de coherencia al aplicar las normas de examen también puede servirnos de ejemplo: es posible que a un alumno/a se le exija que respete a rajatabla el tiempo concedido, mientras que a otro tal vez se le permita un poco de flexibilidad. Por eso, el IB se rige por unas reglas claras y coherentes que se establecen en los *Procedimientos de evaluación* de cada programa. Cuando estas reglas se incumplen, el IB trata la situación como un caso de mala administración.

Sesgos relacionados con la corrección

La mayoría de sesgos relacionados con la corrección son inconscientes, y no deliberados. Por naturaleza, la mente humana utiliza atajos para facilitar la toma de decisiones y, a menudo, esto provoca que haya sesgos inconscientes. Aunque el IB tiene el deber de mitigar estos sesgos, cuando suceden no es culpa de nadie.

El sesgo que se produce al corregir las pruebas puede tener varios motivos, tales como la actitud personal frente a la pulcritud de la escritura del alumnado (por ejemplo, Hughes *et al.*, 1983), el trato preferente en función del género del alumno/a (en caso de que pueda saberlo o sospecharlo el examinador/a), y la indebida atención que se preste a factores como el formato, los signos de puntuación y la ortografía, que pueden no tener una pertinencia significativa en ciertos contextos de evaluación. Para resolver estas cuestiones, los examinadores/as deben recibir una capacitación adecuada y su trabajo debe revisarse.

Hay numerosos estudios acerca de sesgos inconscientes basados en factores como el género, la nacionalidad o el colegio. Para minimizar este tipo de sesgos, el IB procura que el trabajo entregado para evaluación se anonimice antes de enviarlo a corregir.

Otra área de los sesgos que se ha investigado en profundidad es el denominado *efecto halo*. Dicho efecto consiste en que un examinador/a desarrolla una opinión positiva sobre un alumno/a si sus primeras respuestas son de muy buena calidad, lo cual hace que le conceda de manera desproporcionada el beneficio de la duda en respuestas posteriores.

Cuando los examinadores/as reconocen que no pueden corregir de manera justa y sin sesgos una respuesta sobre un tema determinado, el IB cuenta con procesos para que el trabajo lo evalúe alguien que pueda corregirlo sin sesgos, basándose en los méritos académicos del trabajo presentado.

Sesgos relacionados con las tareas de evaluación

Existe sesgo cuando la diferencia en los resultados no está relacionada con lo que se está midiendo. Por ejemplo, usar un ejemplo cultural específico en el cual se basa un problema matemático, en el que la falta de conocimiento sobre una determinada tradición cultural (que no es un requisito de la evaluación) podría dificultar la comprensión y el desempeño (que sí son requisitos de la evaluación).

En la elaboración de las pruebas psicométricas, cuando se demuestra que un elemento o pregunta recibe respuestas inusuales durante las pruebas previas, o que recibe respuestas sustancialmente distintas de diferentes subgrupos del alumnado (lo cual se denomina *funcionamiento diferencial del elemento*), entonces se considera sesgado y se elimina de la prueba. Los subgrupos del alumnado pueden definirse por género, etnia, clase social o competencia lingüística, o por cualquier característica definitoria que pueda considerarse no pertinente con respecto al constructo que se evalúa.

Para determinar si existe sesgo en una evaluación o un elemento de una evaluación, se debe plantear con cuidado cómo se puede vincular explícitamente la tarea con el constructo subyacente y cuáles podrían ser los posibles factores que introduzcan el sesgo. Si la decisión se basa en motivos puramente estadísticos, se corre el riesgo de caer en la trampa de confundir una evaluación sesgada con una evaluación cuyo propósito establece una distinción entre los grupos. Goldstein (1996) y Humphreys (1986) han sugerido que es conveniente distinguir entre *diferencia*, que es un hecho determinado objetivamente, y *sesgo*, que es un juicio sobre la pertinencia de la diferencia.

Black (1999) propone las siguientes seis posibilidades como aquellas que más a menudo pueden hacer que las preguntas sean injustas en cuanto a su impacto en alumnos/as distintos:

1. El contexto en el que se plantea la pregunta puede favorecer al alumnado que esté familiarizado con dicho contexto. Por ejemplo, las referencias culturales a Estados Unidos benefician al alumnado que se encuentra en ese país, y pueden desfavorecer al alumnado que vive en otros sitios.
2. Las preguntas de desarrollo sobre las relaciones humanas pueden favorecer al alumnado de contextos culturales o familiares en los que se fomenta la expresión de las emociones.
3. Las preguntas de opción múltiple pueden favorecer a los chicos.
4. Los trabajos de clase o los trabajos de proyectos pueden favorecer a las chicas.
5. Una pregunta que utilice lenguaje o convenciones asociados a un entorno socioeconómico concreto puede perjudicar al alumnado de un entorno diferente.
6. Algunas preguntas pueden ser inteligibles solamente en ciertos contextos culturales. Por ejemplo, una pregunta sobre personas mayores que viven solas puede resultar bastante ajena en ciertos contextos culturales, pero familiar en otros.

Aparte de esto, aunque existen pruebas de la existencia del sesgo de género, sobre todo en Estados Unidos, no queda claro qué aspectos del formato de las preguntas contribuyen a estos hallazgos.

Al diseñar las evaluaciones, no siempre está claro cómo se deberían abordar las diferencias que hay entre el alumnado que realiza una evaluación concreta. Las evaluaciones deben diseñarse de modo que, mediante diversos tipos de tareas y preguntas, se reduzcan las consecuencias generales del sesgo. Deben evitarse todos los estereotipos culturales o de género, sean explícitos o no. El contenido de las preguntas debe examinarse minuciosamente para evitar que pertenezcan a alguna de las categorías que se sabe que introducen factores de falta de equidad, y la realización de pruebas previas de las preguntas en muestras de diferentes grupos del alumnado pueden revelar casos ocultos de ello. No obstante, si se excluyen todos los tipos de preguntas con sesgo y todos los marcos hipotéticos, quedan pocas opciones disponibles para diseñar evaluaciones y elaborar preguntas. Las limitaciones resultantes tendrán un efecto negativo en la validez de la evaluación. Aparte de evitar los peligros obvios e innecesarios, la solución más razonable parece ser la de un enfoque del diseño de la evaluación equilibrado, utilizando distintos tipos de tareas y formatos de evaluación.

También hay que plantearse cuántos grupos diferenciados del alumnado deben tenerse en cuenta. Por ejemplo, ¿se deberían tener en cuenta a los alumnos/as que tienen distintos estilos de aprendizaje? Según han afirmado Hieronymus y Hoover (1986), si se considera que las diferencias de interés y motivación son factores de sesgo, puede decirse que todas las tareas o métodos de evaluación tienen un cierto grado de sesgo. Por ejemplo, resulta inevitable que los pasajes de texto utilizados en los exámenes de lengua tengan mayor interés para unos alumnos/as que para otros. La equidad en la evaluación, que incluye evitar el sesgo, es un tema de suma importancia, especialmente en ciertos países y regiones o territorios donde cualquier sesgo demostrable en un instrumento de evaluación puede desembocar en un litigio. Sin embargo, demostrar que se trata de sesgo, y no de diferencia en el desempeño, a menudo es una cuestión de interpretación, estrechamente ligada al contexto social concreto en el que se lleva a cabo la evaluación.

Gipps y Murphy (1994) concluyen su libro *A Fair Test? Assessment, Achievement and Equity* diciendo que no existe ni puede existir ninguna prueba que se pueda calificar de justa, ya que la situación es demasiado compleja y la noción demasiado simple. Sin embargo, eso no significa que quienes diseñan las evaluaciones y redactan las preguntas no deban hacer todo lo que esté a su alcance para reducir el efecto del sesgo y de la falta de equidad. Gipps y Murphy también defienden que quienes diseñan las evaluaciones deben tener como objetivo la igualdad de oportunidades y de acceso a la evaluación, en lugar de la igualdad de

resultados que se consigue manipulando elementos de las pruebas según las estadísticas de las respuestas. Estas autoras se preguntan, por ejemplo, en qué medida se justificaría incluir preguntas de opción múltiple en los exámenes de inglés para mejorar el desempeño relativo de los varones, ya que esto distorsionaría la validez de la evaluación según las definiciones ampliamente aceptadas de este término.

Se acepta, en general, que la falta de imparcialidad en el proceso de evaluación es solo uno de los factores que contribuyen a la falta de equidad en la educación, y posiblemente uno de los menos significativos. Hay muchos otros factores causantes de falta de equidad en la educación que tienen un impacto decisivo en los logros del alumnado, por ejemplo:

- Las diferencias de calidad de enseñanza en un mismo colegio
- Las diferencias de nivel de recursos de distintos colegios y en distintas áreas geográficas
- Las diferencias en el nivel de apoyo social y familiar que recibe cada estudiante

Cualquiera de estos factores puede influir de forma significativa en el éxito individual de un alumno/a de un modo tal que ningún proceso de evaluación, por justo que fuese, podría compensar. Smith y Tomlinson (1989), por ejemplo, descubrieron que la eficacia del colegio era un factor mucho más importante para la existencia de diferencias en los resultados de los exámenes que las etnias del alumnado, y señalaron que los intentos de ajustar los instrumentos de evaluación para remediar diferencias en el desempeño por parte de diferentes grupos pueden ser a veces inadecuados.

Las consideraciones de esta naturaleza constituyeron la base lógica para que se realizasen pruebas de aptitud en lugar de pruebas de logro, pero se ha llegado a comprender que no es posible evaluar la aptitud, la capacidad o el potencial puros, separados del origen social y de la experiencia educativa. Tampoco es posible juzgar el logro educativo independientemente del contexto social y cultural. El concepto de éxito en la educación se define y se mide en cualquier sociedad según los estándares de un sector concreto y delimitado de dicha sociedad.

Eliminación de sesgos y de barreras para la evaluación

Otro aspecto del sesgo que debe combatirse es la posibilidad de que una tarea de evaluación discrimine negativa e injustamente a cualquier estudiante. Con este fin, hay que asegurarse de que las condiciones de evaluación permitan aplicar las adecuaciones necesarias para eliminar las barreras para la evaluación. Así se garantiza que todo el alumnado puede demostrar su nivel de logro educativo en condiciones de igualdad con respecto al resto.

Esta idea se trata con más detalle en la sección [“Priorización de la equidad para todo el alumnado”](#) de esta publicación.

Cómo reconocer el sesgo

Es importante tener en cuenta que los sesgos pueden ser positivos y también negativos. Si un grupo de estudiantes está especialmente familiarizado con una tarea, o le resulta más sencillo completarla, esto se considera sesgo. El objetivo de una evaluación justa es ofrecer las mismas oportunidades a todo el alumnado.

En esta publicación ya se ha descrito de qué manera se pueden introducir sesgos en el diseño de las evaluaciones o en el proceso de corrección. Si bien es muy importante tener una actitud proactiva al pensar en los sesgos durante esas partes del ciclo de evaluación, no basta con pensar en los posibles sesgos que podría haber. También es necesario buscar pruebas de dichos sesgos en los resultados del alumnado, para lo cual hay que comparar los resultados en una pregunta concreta con el resto del examen en conjunto.

Asimismo, también es importante que las decisiones relacionadas con los sesgos se basen en pruebas y no en estereotipos.

Equiparabilidad

La equiparabilidad es uno de los aspectos más complejos de la validez. A menudo los resultados de la evaluación se utilizan para comparar a estudiantes con fines de selección. Si dos alumnos/as han realizado el mismo examen al mismo tiempo, se puede afirmar con certeza que un alumno/a con una calificación

de 7 ha tenido mejor desempeño ese día que un alumno/a que haya logrado una calificación de 4. Sin embargo, con frecuencia se realizan comparaciones más complejas que esa, como se puede ver en los siguientes ejemplos:

- Dos estudiantes obtuvieron una calificación de 6 en Historia, pero respondieron a preguntas distintas o eligieron opciones distintas.
- Dos alumnas, una que obtuvo una calificación de 5 en Literatura española en mayo de 2025 y otra que alcanzó una calificación de 5 en la misma asignatura en noviembre de 2023.
- Dos alumnos, uno que obtuvo una calificación de 4 en Física y el otro una calificación de 4 en Química.
- Dos alumnas, una que obtuvo una calificación de 4 en Matemáticas y la otra una calificación de 4 en Geografía.
- Un alumno de una ubicación geográfica con una calificación de 3 en Informática y una de 6 en Literatura china, y una alumna de otra ubicación geográfica con una calificación de 5 en Literatura japonesa y una de 4 en Biología.
- Dos alumnas de 15 años, una que obtuvo un certificado del PAI y otra que obtuvo una titulación de otra organización educativa.
- Dos alumnos, uno de los cuales obtuvo una calificación de 6 en Indonesio B de Nivel Medio (NM) y otro que alcanzó una calificación de 5 en Indonesio B de Nivel Superior (NS).

La equiparabilidad se pregunta si, de alguna manera, dos resultados de evaluación distintos se pueden considerar como iguales. Esto resulta especialmente difícil entre asignaturas, ya que primero el IB evalúa cosas distintas y luego se plantea la pregunta de si tienen un valor equivalente.

Aquí resulta especialmente pertinente el concepto de validez para un propósito concreto, en vez de como característica de una evaluación en abstracto. El resultado que obtenga un alumno/a en la asignatura de Música es un mejor indicador para conocer su predisposición a dedicarse a la música de manera profesional que la misma calificación en la asignatura de Artes Visuales, pero ambos resultados pueden ser equivalentes si lo que se quiere predecir es su preparación para estudiar Historia.

El tema de la equiparabilidad se complica aún más, ya que cada estudiante tiene sus puntos fuertes y sus puntos débiles, por lo que tendrá más facilidad en unas asignaturas que en otras. Esto supone un desafío particular para el IB, ya que los grupos de estudiantes que realizan cada examen no son iguales, sobre todo allí donde hay disponibles varios cursos.

El IB procura mantener tres principios acerca de la equiparabilidad:

- La calidad de trabajo necesaria para lograr las distintas calificaciones en una asignatura o disciplina debe ser equiparable entre convocatorias de un año y entre años distintos.
- Las calificaciones finales deben tener el mismo significado en todas las asignaturas o disciplinas, de tal modo que los distintos itinerarios para lograr la titulación del programa (por ejemplo, el diploma del IB o el certificado del PAI) sean equiparables.
- Si bien el IB mantiene su compromiso con la evaluación de las habilidades de pensamiento de orden superior, también se presta atención a garantizar que las calificaciones obtenidas en otras titulaciones sigan siendo equiparables a las obtenidas en el IB.

Cómo medir la equiparabilidad

Hay muchas maneras de estimar en qué medida son equiparables dos evaluaciones, y este tema ha sido objeto de numerosos artículos académicos. En su revisión de la bibliografía especializada sobre el tema de la equiparabilidad entre asignaturas, Coe *et al.* (2008) estudian los métodos que existen para comparar la dificultad de los exámenes de diferentes asignaturas y los dividen en dos grandes categorías: los métodos estadísticos y los métodos estimativos.

Los métodos estadísticos se centran en comparar el desempeño del alumnado en las evaluaciones, con el fin de encontrar patrones. Estos métodos se basan en la idea de que, si dos evaluaciones son equiparables, entonces en una muestra aleatoria lo suficientemente grande de estudiantes que realizan las dos, los resultados por lo general deberían ser los mismos en ambas evaluaciones.

Por el contrario, los métodos estimativos utilizan a personas expertas en la asignatura para estudiar varias evaluaciones y dar su opinión fundamentada acerca de la dificultad relativa de cada una. En estos casos, se utiliza una variedad de herramientas y técnicas de investigación para garantizar que la comparación se realice entre elementos similares.

Ambos enfoques presentan deficiencias conceptuales graves, y Coe *et al.* (2008) identifican seis críticas amplias para cada técnica:

Tabla 4

Críticas a los métodos estadísticos y a los métodos estimativos para comparar evaluaciones

Críticas a los métodos estadísticos	Críticas a los métodos estimativos
<ul style="list-style-type: none"> • Los estudios pueden medir otros factores aparte de la dificultad, como son la enseñanza o la motivación. • Multidimensionalidad: es posible que las asignaturas no tengan ninguna característica en común. • Falta de representatividad: ¿las estadísticas se basan en un grupo de estudiantes con algún sesgo inherente? • Diferencias entre subgrupos: si distintos grupos de estudiantes se enfrentan a distintos niveles de dificultad, ¿esto pone en entredicho las conclusiones de la equiparabilidad relativa? • Desacuerdo entre los métodos: ¿es posible que un método sea el “correcto”? • Problemas de forzar la igualdad: ¿cómo afectaría al alumnado que se enfrenta a la evaluación? 	<ul style="list-style-type: none"> • Amplitud de criterios: para poder aplicarlos en distintas asignaturas, los criterios deben ser muy amplios, y, por tanto, imprecisos. • Valoración de las respuestas en distintos niveles de exigencia: el personal examinador tiende a valorar mejor una buena respuesta a una pregunta fácil que una respuesta más débil a una pregunta más difícil. • Valoración de tipos distintos de trabajo: el personal examinador tiene dificultades para comparar tipos distintos de tareas, por ejemplo, respuestas cortas y ensayos. • Empleo de comparaciones estadísticas: incluso los métodos “de juicio” se basan en la experiencia de cómo es probable que sea el desempeño del alumnado típico. • Interpretación y contexto: las evaluaciones estimativas de la capacidad o la equidad del alumnado deben tener en cuenta diferencias estructurales, como las que existen entre los exámenes finales únicos y una serie de evaluaciones modulares. • Suma de varios juicios: la mayoría de las evaluaciones miden varios criterios que se deben equilibrar.

Quienes defienden una de estas metodologías suelen criticar bastante la otra. Además, por otro lado, hay quienes argumentan que todos los enfoques actuales tienen defectos de base. A fin de investigar este tema, Coe *et al.* (2008) aplicaron cinco maneras distintas de medir la equiparabilidad entre asignaturas de dos titulaciones de Inglaterra: el GCSE (General Certificate of Secondary Education) y el A-Level (Advanced Level). La conclusión fue que hay un nivel razonablemente elevado de coherencia entre las maneras estadísticas y las estimativas de medir la equiparabilidad, y que las diferencias entre ellas eran mucho menores que las diferencias entre asignaturas. Además, también concluyeron que las dificultades relativas de las asignaturas se mantuvieron estables de un año a otro.

El IB mantiene la equiparabilidad entre los años y las opciones a través de la triangulación de las valoraciones de los examinadores/as, el análisis estadístico, y los comentarios del profesorado (véase la sección “[Concesión de calificaciones finales y totalización](#)”). La equiparabilidad también se revisa a nivel de asignaturas o disciplinas mediante el uso de enfoques estadísticos y de juicios de personal experto.

El enfoque del IB acerca de la validez

- El IB considera que las evaluaciones auténticas y pertinentes al constructo son más importantes que maximizar la fiabilidad.
- El IB cree en una educación holística y completa. Su prioridad es contar con argumentos sólidos que muestren la validez de cada programa. Esta validez es más importante que la de los cursos individuales o de los distintos itinerarios opcionales de los cursos.

La validez consiste en una búsqueda compleja y polifacética de un equilibrio entre diversas demandas importantes y opuestas. No existe una única forma correcta de equilibrar los elementos contrapuestos de la validez: en última instancia, el punto de equilibrio que se elige es una decisión que se basa en los valores de la organización que desarrolla las evaluaciones.

Para el IB es fundamental que las evaluaciones aborden **aquello que es importante**, y que lo hagan de una manera que **refleje el mundo real**. El primero de estos aspectos está englobado en el concepto de *pertinencia del constructo*. Esto significa que en las evaluaciones del IB no se pregunta simplemente lo que resulta fácil de corregir, sino aquello que es realmente importante en cada asignatura. El segundo de estos aspectos hace referencia a la *autenticidad*, que significa que las evaluaciones del IB las componen tareas significativas que presentan situaciones que el alumnado puede encontrar en el mundo real, y no tareas artificiales y superficiales.

Estos objetivos tienen prioridad sobre otros aspectos de la validez, principalmente la **fiabilidad** de la evaluación. Por lo general, las tareas auténticas y significativas requieren un alto grado de subjetividad en la corrección, lo que significa que hay que aceptar mayores variaciones entre los examinadores/as que si, por ejemplo, se utilizasen evaluaciones con preguntas de opción múltiple que tienen una respuesta clara y se pueden corregir de manera objetiva. Además, estas tareas también afectan a la **manejabilidad** de la evaluación. Los tipos de evaluaciones que prefiere el IB son más complicados y requieren más tiempo para su creación, administración y corrección. Además, estas evaluaciones exigen un mayor compromiso por parte del alumnado. Esto provoca que aumente su carga de trabajo, ya que, por ejemplo, debe realizar tareas de investigación concienzudas en vez de centrarse solamente en su desempeño en los exámenes.

Si bien es aceptable que otros organismos tengan distintas prioridades a la hora de buscar el equilibrio en la validez de la evaluación, el IB está convencido de que la postura de la organización resulta adecuada y justificable en el sistema que utiliza de verificación externa de las evaluaciones.

El objetivo del IB es lograr más que otros currículos, al formar personas solidarias, informadas y ávidas de conocimiento motivadas para tener éxito. Como indica en su declaración de principios, el IB, mediante sus programas educativos, espera que sus estudiantes sean capaces de contribuir a crear un mundo mejor, en el marco del entendimiento mutuo y el respeto intercultural. Es más, todos los programas del IB están comprometidos con el desarrollo del alumnado conforme al perfil de la comunidad de aprendizaje del IB, que identifica diez atributos que el IB valora y busca desarrollar en sus estudiantes.

En lo que concierne a la evaluación, esto significa que los propósitos del IB se definen a nivel de programa, no al nivel de cada asignatura o disciplina. Por lo tanto, la cuestión de la validez de la evaluación se debe plantear a nivel de programa y debe incluir las normas que rigen la concesión de la titulación completa del programa correspondiente, y no solo de cada una de las calificaciones individuales.

Sin embargo, esto no significa que no sea importante tener en cuenta cada curso o cada tarea de evaluación individual; de hecho, algunos aspectos de la validez solamente tienen sentido a este nivel de detalle. No obstante, al realizar un argumento de validez general, el IB tiene en cuenta el programa de estudios completo que ha cursado el alumno/a.

La mayoría de principios de evaluación que se presentan en este recurso son aplicables a toda la educación del IB, pero en la [“Sección C: Procesos específicos de los programas del IB”](#) se aborda cada programa individualmente y se tratan sus peculiaridades.

Beneficios de la evaluación digital para la validez

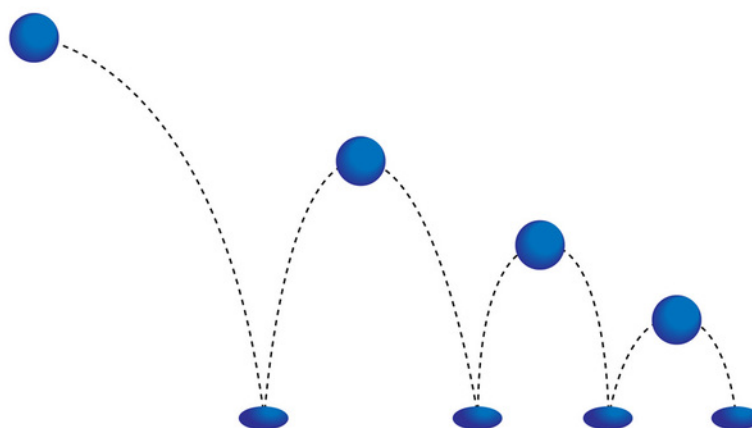
- En el mundo actual, el papel ya no es el soporte preponderante: los dispositivos informáticos forman parte de todas las facetas de la vida.
- Los exámenes en papel no pueden ofrecer imágenes en movimiento ni permiten que haya una interacción significativa con el alumnado.
- La versatilidad de los dispositivos permite a cada estudiante hacer las modificaciones visuales y sonoras que necesite para acceder a las preguntas, en vez de tener que solicitar pruebas de examen modificadas con varios meses de antelación.
- A pesar de las preocupaciones que puedan surgir, la evaluación digital permite protegerse mejor contra las conductas impropiedades del alumnado y la mala administración por parte de los colegios.

Los correos electrónicos, los mensajes de texto y las redes sociales son actualmente mucho más frecuentes que las cartas como medio de comunicación. Asimismo, hoy en día muchos empleos requieren utilizar dispositivos informáticos. Por lógica, ofrecer evaluaciones auténticas pasa por incluir en ellas el uso de la tecnología y los dispositivos.

Tradicionalmente, las evaluaciones han resultado limitadas en lo referido a aquello que pueden cubrir con medios tradicionales (por ejemplo, el uso de lápiz y papel). En los exámenes escritos, el alumnado se ha visto limitado a responder a un estímulo sencillo o a una pregunta simple sin tener oportunidad de manipulación o de interacción. En el nivel más básico, las evaluaciones digitales ofrecen la posibilidad de incluir como estímulo elementos de audio y video a los que cada estudiante puede acceder, en vez de tener que depender de un supervisor/a de examen que se encargue de reproducir el material para toda la clase. En el Programa de los Años Intermedios (PAI), la evaluación digital también permite utilizar diagramas animados y simulaciones donde resulte adecuado, en vez de incluir una imagen estática y una descripción larga sobre el movimiento de un objeto.

Figura 13

Mejora de la claridad de una pregunta de evaluación electrónica mediante el uso de una imagen en movimiento



A medida que las tareas de evaluación evolucionan, la tecnología hace posible evaluar de qué manera el alumno/a aborda un problema, responde ante información nueva o interactúa con una simulación. Con los sistemas de tutoría inteligente, la evaluación digital debería permitir responder a las reacciones de cada estudiante. Así se recrearían las oportunidades que ofrecen las evaluaciones que realiza el profesorado, como los exámenes orales, pero sin los problemas que provoca el hecho de que distintos alumnos/as tengan docentes diferentes. Como resultado, un beneficio clave de la evaluación digital es que permite diseñar nuevas tareas para evaluar lo que el IB realmente quiere medir, sin limitarse a lo que se puede juzgar mediante un examen en papel.

Otro aspecto clave de la validez es reducir los sesgos, sobre todo cuando se necesitan adecuaciones inclusivas de acceso. Periódicamente, el IB procesa solicitudes para realizar exámenes con distintas fuentes o en papel de otro color. Asimismo, también recibe solicitudes para preparar cuestionarios de examen en formato braille.

Si bien las evaluaciones digitales no pueden eliminar por completo estas barreras, la tecnología puede permitir al alumnado seleccionar el tamaño de letra y el color que mejor se adapten a sus necesidades o preferencias. El empleo de software que permite el uso de lectores de pantalla también ofrece otros beneficios destacables en lo referido a la accesibilidad. Así, aunque es posible que parte del alumnado no pueda utilizar sus dispositivos para cubrir sus necesidades específicas de acceso, la evaluación digital permite que una cantidad mayor de estudiantes pueda acceder a las pruebas sin necesitar apoyo adicional, si bien este apoyo se podrá seguir utilizando cuando resulte adecuado.

Una de las preocupaciones más destacadas sobre la evaluación digital —que también supone una importante amenaza para su validez— es la seguridad de las pruebas. Algunas perspectivas críticas a la evaluación digital afirman que es menos segura que los exámenes en papel. En realidad, las evaluaciones digitales están protegidas frente a algunos riesgos de los exámenes en papel, si bien se exponen a otros.

Una de las ventajas más significativas de los exámenes digitales es que se entregan de forma segura y solo se puede acceder a ellos dentro de un plazo definido por el IB, sin que los usuarios puedan acceder al contenido del examen antes de que el colegio les conceda acceso. Si esta situación se compara con la de enviar los exámenes en papel —que implica que cualquier persona puede leerlos y que los colegios deben guardarlos a salvo durante días antes del examen—, la evaluación digital resulta más ventajosa.

El IB es consciente de que la evaluación digital presenta otros desafíos durante el examen en sí, ya que el alumnado tiene acceso a dispositivos que se conectan a Internet. Para reducir el riesgo de que los alumnos/as accedan a material no autorizado a través de sus dispositivos, los exámenes digitales del IB funcionan en modo restringido, lo cual impide el acceso a cualquier otro contenido mientras se están realizando, con el fin de garantizar la seguridad de los exámenes y la integridad académica. Además, la capacidad de registrar cómo responde a las preguntas el alumnado (por ejemplo, archivos de registro), así como sus respuestas finales, ofrece oportunidades para identificar comportamientos relacionados con conductas improcedentes.

Estándares

- Cualquier definición de estándares comprende tres partes: el **currículo**, la **evaluación**, y el **desempeño**.
- La evaluación formativa normalmente se centra en los estándares curriculares y de evaluación.
- El IB utiliza un **enfoque basado en logros** (anteriormente denominado *enfoque flexible basado en criterios*); esto significa que, para fijar el estándar, emplea de manera equilibrada los criterios de evaluación y una comparación con los resultados del alumnado de años anteriores.
- Mantener los estándares resulta tan importante como fijarlos.

Los tres aspectos de los estándares

En el contexto de la evaluación, la palabra *estándares* se refiere normalmente a la dificultad de las tareas que se encomiendan al alumnado. Los estándares son un componente central de la equiparabilidad. El concepto de estándar se puede considerar de tres maneras distintas:



Cada uno de estos tres aspectos puede variar para cambiar la dificultad general de la asignatura. Sin embargo, los cambios pueden tener distintos efectos y también afectar a los plazos.

Para el IB, los cambios **curriculares** normalmente se producen durante la revisión del currículo, aunque es posible realizarlos fuera del ciclo de revisión formal. La equidad exige que el IB otorgue al profesorado el tiempo suficiente para adaptar su enseñanza a los cambios que se produzcan. En el caso del Programa del Diploma (PD) y el Programa de Orientación Profesional (POP), el plazo ideal serían dos años (el tiempo que dura el curso). Para el Programa de los Años Intermedios (PAI) podría hacer falta más tiempo, ya que el programa dura cinco años, aunque normalmente las evaluaciones solo se efectúan en los dos últimos años.

Los estándares de **evaluación** dependen del ciclo de desarrollo de exámenes, que normalmente tiene una duración de entre 12 y 18 meses. A diferencia de los estándares curriculares, los estándares de evaluación

varían de un año a otro, ya que es imposible diseñar dos grupos de preguntas que tengan exactamente el mismo nivel de dificultad. Incluso repetir las mismas preguntas que en el examen anterior supondría un desafío menor para el alumnado, ya que, en cierta medida, estaría familiarizado con esas preguntas la segunda vez. Es posible que en ocasiones se tome la decisión de cambiar los estándares de evaluación de un examen si los exámenes anteriores no obtuvieron los resultados esperados. En función de la dimensión del cambio que haga falta, también es posible modificar un examen que ya esté en proceso de elaboración. También en estos casos la equidad exige que el IB notifique al profesorado los cambios realizados, pero, dependiendo de la profundidad de dichos cambios, dicha notificación se haría con mucha menos antelación que en el caso de los cambios curriculares.

Los estándares de **desempeño** (es decir, los límites de calificación) se ajustan cada año para equilibrar los cambios que haya podido haber en los estándares de evaluación. Por ejemplo, si el examen resulta más difícil y por ese motivo el alumnado en general obtiene puntuaciones más bajas, se bajarían los límites de calificación para que el alumnado con un desempeño igual al del año anterior obtenga la misma calificación que habría logrado entonces. A diferencia de los otros dos tipos de estándares, es sobre todo el personal examinador, más que el profesorado, quien debe comprender los estándares de desempeño, con el fin de poder mantenerlos en las distintas convocatorias de exámenes. Es posible cambiar el estándar de desempeño en el momento en el que se establecen los límites de calificación. El IB normalmente avisaría con antelación a los colegios para que pudieran gestionar y ajustar lo que se espera del alumnado. En tal caso, el paso más importante es que el personal examinador comprenda cómo es el nuevo estándar para así poder aplicarlo a las futuras convocatorias de exámenes.

Al establecer un estándar general para una evaluación, es importante encontrar un equilibrio entre estas tres definiciones. Aunque es posible preparar preguntas extremadamente difíciles a partir de un material simple, o fijar estándares de desempeño extremadamente altos a partir de una serie de preguntas sencillas, esto suele provocar que haya niveles muy bajos de pertinencia del constructo. Por ejemplo, requerir la máxima puntuación en una prueba para alcanzar una calificación de 7 exige al alumnado demostrar niveles de precisión muy altos. Así, un alumno/a con una comprensión excelente de la asignatura, pero que preste poca atención a la redacción o que tenga pocas habilidades de expresión escrita difícilmente obtendrá una calificación de 7. ¿Es este el propósito pensado para la evaluación?

La definición de *estándares* se aplica tanto a las evaluaciones **formativas** como a las **sumativas**. Sin embargo, en la evaluación formativa, los estándares de desempeño normalmente quedan a juicio de cada docente según lo que quiera que se aprenda. Normalmente, en la evaluación formativa se suele prestar más atención a los estándares curriculares y de evaluación, con el fin de que las pruebas resulten adecuadas para el alumnado y que proporcionen información útil para fundamentar la enseñanza en el futuro.

El enfoque normativo y el enfoque basado en criterios

- El *enfoque normativo* implica clasificar en orden de calidad el desempeño del alumnado y otorgar cada calificación a una proporción fija de estudiantes.
- El *enfoque basado en criterios* implica fijar el estándar de desempeño en función de una descripción de lo que se busca en el desempeño del alumnado.
- El IB utiliza un *enfoque basado en logros* (también denominado *enfoque flexible basado en criterios*); esto significa que, para fijar el estándar, emplea de manera equilibrada los criterios de evaluación y una comparación con los resultados del alumnado de años anteriores.

El enfoque normativo y el enfoque basado en criterios representan dos maneras de fijar y mantener el estándar de desempeño en las evaluaciones.

Enfoque normativo

El enfoque normativo se suele relacionar con pruebas estandarizadas. El principio consiste en aplicar una prueba piloto a una muestra típica de estudiantes, y después utilizar los resultados (a los que, por

definición, debe corresponder una distribución normal o curva campaniforme) como escala de referencia para la puntuación de cualquier estudiante que haga la misma prueba después. El proceso de derivar una distribución estándar de las puntuaciones a partir de la prueba inicial se llama *normalización*.

Esta descripción técnica de *enfoque normativo* no implica necesariamente aplicar una distribución fija a cada conjunto de resultados de la prueba: la distribución fija solo se utiliza en la normalización inicial. La distribución de las puntuaciones de posteriores estudiantes puede ser distinta de la distribución normal.

En la práctica, se suele usar el término *enfoque normativo* para referirse al proceso en el que se coloca a al alumnado por orden según su desempeño y a establecer la proporción de estudiantes que recibirá cada calificación. Por ejemplo, el alumnado incluido en el mejor 15 % de los resultados recibirá la calificación más alta.

Enfoque basado en criterios

Glaser (1963) fue el primero que propuso la noción de *evaluación basada en criterios*. Dicha evaluación supuso un cambio de dirección en el ámbito de establecer estándares de desempeño, ya que dirigió el énfasis hacia medir los logros del alumnado con respecto a un ámbito de conducta bien definido (Popham, 1978).

En el enfoque basado en criterios, el desempeño de cada estudiante se compara con una descripción predefinida de lo que se espera en cada calificación. Normalmente, esta función la desempeña personal experto de la asignatura o de la evaluación mediante el uso de su juicio profesional.

Una limitación de este enfoque es que resulta muy complicado crear descripciones inequívocas y que signifiquen lo mismo para todo el personal experto. De hecho, se ha argumentado que “ningún criterio, independientemente de lo bien formulado que esté, permite una interpretación inequívoca” (William, 1993). El resultado de una prueba tradicional de evaluación basada en criterios es que se demuestra o no el dominio del área pertinente.

En la práctica, ambos enfoques presentan importantes desventajas. Un enfoque normativo estricto exige demostrar que la prueba que se realiza tiene el mismo nivel de dificultad que la prueba inicial, mientras que el enfoque basado en criterios está sujeto al efecto Good y Cresswell (1988), según el cual el juicio del personal experto no tiene en cuenta con precisión la exigencia de las preguntas.

Enfoque basado en logros

El IB utiliza un enfoque basado en logros (Newton, 2011), anteriormente denominado enfoque flexible basado en criterios (Baird *et al.*, 2000), el cual se basa en criterios y también reconoce el efecto Good y Cresswell (1988). En términos prácticos en el IB, se pide al personal examinador experto que, partiendo de criterios (descriptores de calificaciones finales), establezca un rango estrecho sobre el cual basar los límites de calificación. Posteriormente, se realiza una comparación con los límites calculados de tal modo que coincidan con el desempeño de años anteriores. Las calificaciones finales se fijan en los puntos en los que estos dos límites coinciden. En caso de que no coincidan, se llevan a cabo discusiones posteriores sobre cómo armonizarlos.

Finalmente, es importante recordar que las pruebas de evaluación, ya procedan del enfoque normativo o del enfoque basado en criterios, se diferencian en mayor medida por el análisis y la interpretación de las respuestas del alumnado que por el tipo de preguntas planteadas.

Mantenimiento de los estándares

- Una vez establecidos los estándares adecuados, el IB tiene que asegurarse de que se aplican cada año.
- Los estándares curriculares se reconsideran durante la revisión del currículo.
- Los estándares de evaluación y los de desempeño se mantienen mediante una combinación de juicio profesional y pruebas estadísticas.

La equiparabilidad es un aspecto esencial de la validez. En el contexto del mantenimiento de los estándares, esto significa que el IB debe asegurarse de aplicar los mismos estándares cada año. Los estándares del IB se basan en el significado de las calificaciones y no de las puntuaciones. La diferencia entre ambos conceptos se explica en la sección “[Diferencia entre puntuaciones y calificaciones](#)”.

Los **estándares curriculares** son los más fáciles de mantener, ya que no cambian de una convocatoria de exámenes a otra. Las evaluaciones deben ser un fiel reflejo de todo el currículo, y esto se supervisa durante la elaboración de los exámenes. Sin embargo, hay factores externos que pueden provocar que se den cambios en los estándares curriculares. Las habilidades informáticas ilustran perfectamente esta idea. Los dispositivos van evolucionando y cada vez son algo más común en el día a día. Así, los conocimientos y la comprensión que una vez estuvieron solo al alcance de especialistas por su complejidad, ahora son conocimientos cotidianos. Por lo tanto, el estándar relativo al currículo para este tema ha cambiado, aunque el contenido siga siendo el mismo.

Para resolver problemas de este tipo y para mantener los contenidos actualizados, el IB se sirve de un ciclo de revisiones curriculares. A menudo, los estándares curriculares cambian como resultado de estas revisiones y, en consecuencia, el IB debe equilibrar los estándares de evaluación y los de desempeño para que el alumnado no quede en situación de desventaja.

Los **estándares de evaluación** en los exámenes del IB cambian en cada convocatoria de exámenes, ya que es imposible que dos exámenes tengan exactamente el mismo nivel de exigencia. En algunos sistemas educativos se realizan numerosas pruebas previas para determinar la exigencia de cada pregunta, y los exámenes se preparan minuciosamente para garantizar un nivel de dificultad adecuado. En el IB no se realizan pruebas previas de los exámenes, debido al riesgo de que las preguntas se divulguen antes de que el alumnado haga sus exámenes. Por tanto, se confía en el juicio profesional y la experiencia de los autores/as de exámenes y de los verificadores/as para crear exámenes coherentes y equilibrados.

El IB también ajusta los **estándares de desempeño** en función de las respuestas del alumnado para que se mantenga el estándar general. En la sección anterior se ha abordado cuáles son las dos formas principales de mantener los estándares de desempeño. Precisamente mantener los estándares de desempeño es el propósito del proceso de concesión de calificaciones del IB.

Descripción del éxito del alumnado en la evaluación sumativa

- Si bien las calificaciones representan una visión muy simplificada de los logros del alumnado, permiten que otras partes interesadas (como las universidades, las empresas o los institutos de educación superior) dispongan de una base sobre la que realizar juicios razonables de selección.
- Si se proporcionara información más compleja y holística, entonces otras partes interesadas deberían simplificar dicha información para tomar decisiones sensatas de selección, y es probable que no fuesen tan cuidadosas como lo es el IB al establecer los límites de calificación.
- En las evaluaciones del IB, el juicio profesional es importante para lograr resultados significativos para el alumnado. El IB es consciente de que debe respaldar ese juicio con pruebas objetivas y sin sesgos con el fin de garantizar la equidad.

El impacto de las calificaciones

Pensemos un momento en los conocimientos, las habilidades y la experiencia de un chef excelente. Este chef tiene conocimientos: sobre los ingredientes y las combinaciones de sabores que darán lugar a un plato perfectamente equilibrado. También tiene habilidades: para seleccionar y preparar los ingredientes, para cocinar, y para emplatar y presentar los platos. Finalmente, tiene experiencia: con utilizar las técnicas para alcanzar la perfección, con juzgar cuándo algo está estupendamente cocinado, y con cómo presentar una combinación de ingredientes y platos que produzcan una gran satisfacción en la mesa.

Dicho todo esto, ahora hay que reducir toda esa complejidad a una calificación de 1 a 7 para decidir quién es el mejor chef. El resultado obtenido apenas significa nada. Aumentar la escala de calificación hasta 100, o incluso hasta 1.000, seguramente no serviría de mucho. Esto permitiría una mayor variedad de calificaciones para diferenciar entre chefs, pero el problema fundamental de intentar comparar conjuntos de habilidades distintos sigue presente.

Eso es exactamente lo que pasa en la evaluación: ¿cuál es la mejor manera de representar en un único resultado la complejidad del conocimiento, la comprensión y las capacidades del alumno/a? Incluso aunque se pudiera obtener y resumir con precisión toda la información sobre el alumno/a, el IB seguiría sin poder concederle una calificación que refleje perfectamente sus talentos.

Otra sugerencia que se suele proponer es no conceder ninguna calificación. Tal vez el IB podría darle a cada estudiante una descripción personalizada de las cosas que hizo bien y de los aspectos en que necesita mejorar. Este enfoque es coherente con los principios de un buen aprendizaje y una buena enseñanza, pero tiene la gran desventaja de que es muy difícil hacer comparaciones entre estudiantes.

El propósito de la evaluación del IB es ayudar al alumnado a demostrar sus logros y progresos para que pueda avanzar en sus estudios o incorporarse al mercado laboral. Esto significa que es necesario algún tipo de proceso de selección por parte de las instituciones que reciben al alumnado (normalmente universidades en el caso del PD o el POP).

Claramente, si los otros criterios son menos fiables que los exámenes, confiar más en dichos criterios provocará que se tomen decisiones de selección menos fiables.

(Cresswell, 1986, p. 42)

Si va a haber un proceso de selección, el IB tiene la responsabilidad de apoyar a sus estudiantes para que las decisiones de selección sean lo más justas y tengan el máximo sentido posible. Si las instituciones receptoras solo reciben informes descriptivos, entonces se encargarán de buscar algún otro modo de comparar al alumnado, y casi con total seguridad los sistemas que utilicen serán menos fiables y equiparables que los resultados de los exámenes expresados en calificaciones.

Esto no significa que las evaluaciones sumativas por sí solas sean perfectas, y ni siquiera tal vez una manera especialmente buena de seleccionar estudiantes, pero sí son una forma más justa que muchas de las alternativas. Además, cuando se utilizan junto con otros indicadores pertinentes (como el promedio de calificaciones, las actividades extracurriculares, los ensayos personales, y las cartas de recomendación), pueden proporcionar información útil para la toma de decisiones. Aún más importante, el IB vela constantemente para que sean lo más justas, significativas y fiables posible.

Figura 15

¿Deberíamos hacer distinciones entre ambos estudiantes de cada pareja?



En el ejemplo de la figura 15, la primera pareja de estudiantes se diferencian mediante el uso de los descriptores de calificaciones finales genéricos del PD para una calificación de 5 y una de 6. Si se tuviera que elegir entre los dos alumnos/as, seguramente la decisión sería justa, y es probable que, si volvieran a realizar los exámenes, obtendrían los mismos resultados. En la segunda pareja de estudiantes, probablemente la diferencia no resultará significativa, y es probable que ambos alumnos/as tengan la misma capacidad.

Este ejemplo demuestra que no todas las diferencias son significativas y que el uso de calificaciones aporta una indicación de los aspectos que pueden permitir diferenciar entre dos estudiantes. Es cierto que para un alumno/a que esté en el límite entre dos calificaciones (ya sea un poco por encima o un poco por debajo), cualquiera de las dos será justa, pero, para la mayoría del alumnado, la diferencia de calificaciones indica una diferencia significativa en el desempeño general.

Esto nos lleva a la cuestión de cuántas calificaciones habría que tener. Si solo existen dos calificaciones (aprobado o no aprobado), la mayoría del alumnado obtendrá una calificación justa, pero quienes estén rozando el límite entre ambas sufrirán consecuencias muy graves. Por el contrario, si hubiera 20 calificaciones distintas, habría una cantidad mucho mayor de estudiantes en un límite de calificación o cercano a él, si bien las consecuencias en este caso serían menos graves para cada estudiante. Cresswell (1986) explora este concepto con más detalle.

En general, el IB ha elegido **siete calificaciones** para representar la cantidad de categorías importantes derivadas de sus evaluaciones, y también para encontrar un equilibrio justo entre el número de estudiantes que se encuentren en un límite de calificación y las consecuencias que tendría recibir una calificación equivocada.

La importancia del juicio profesional

Las habilidades de pensamiento complejas y de nivel superior en las que se centra la evaluación del IB no se prestan fácilmente a una corrección sencilla y exenta de juicios. Es probable que las respuestas del alumnado sean muy variadas y adopten muchas formas igualmente correctas y válidas. Las investigaciones realizadas indican que los conocimientos y habilidades complejos no se deberían reducir a bloques pequeños y concretos para enseñarse, y este mismo principio se aplica a la corrección de esas respuestas. Al desarrollar los esquemas de calificación, el personal experto en las asignaturas del IB debe dar una orientación muy precisa sobre cómo corregir el trabajo de todo el alumnado de manera coherente.

Esto implica hacer un gran hincapié en el juicio profesional del personal examinador y, sobre todo, en la pericia profesional del equipo examinador supervisor, que es quien fija y explica los estándares de corrección. En conjunto, esto representa un gran desafío para la fiabilidad del sistema de evaluación, pero es un desafío que hay que superar en el contexto de la precisión de los resultados.

Corrección de las evaluaciones del IB

- La corrección es el proceso de evaluar si el alumno/a ha completado bien la tarea que se le había asignado.
- La corrección se puede centrar en los detalles individuales de cada respuesta o adoptar una posición holística.
- En la evaluación formativa, la corrección puede no conllevar una puntuación numérica, sino ser puramente descriptiva.

Definición de corrección

En esta sección, los términos *corrección* y *métodos de corrección* se refieren a la corrección validada por el IB de las evaluaciones sumativas en el PAI, el PD y el POP. La evaluación formativa se utiliza en todos los programas, pero en muchos casos se realiza en el aula y el IB no valida las evaluaciones mediante la corrección de un examinador/a.

La puntuación obtenida en la corrección no describe lo bien que lo ha hecho un alumno/a. El nivel de logro del alumno/a depende de aspectos como la dificultad de las preguntas y cuál es la puntuación esperada para obtener un aprobado en cada nivel. En la corrección se compara la respuesta de un alumno/a con lo que se espera que sea una respuesta perfecta.

La corrección se puede realizar de varias maneras en función de la naturaleza de la tarea y del tipo de corrección que el instrumento de evaluación determine como apropiado.

Métodos de corrección

El IB utiliza diversos instrumentos de evaluación, y los requisitos de cada uno de ellos ayudan a determinar qué tipo de corrección es la más adecuada.

A veces la corrección es muy **objetiva**: la respuesta del alumno/a es correcta o no lo es. Esto suele pasar cuando la pregunta solo requiere un par de palabras clave o seleccionar una respuesta entre varias posibles. La corrección automática es el proceso de utilizar la tecnología para evaluar el trabajo del alumnado según un esquema de calificación predefinido en el que una respuesta es objetivamente correcta o incorrecta. El IB utiliza la corrección automática para las preguntas de opción múltiple en las que las respuestas son definitivamente correctas o incorrectas, y no requieren de la valoración del examinador/a para estar seguro de que una respuesta es 100 % correcta. El IB no utiliza la corrección asistida por inteligencia artificial en lugar de personas cuando hay una variedad limitada de respuestas esperadas pero se requiere una valoración.

En otras ocasiones, la corrección es mucho más **subjectiva**. En esos casos, puede ser necesario que un examinador/a valore de manera holística si la respuesta del alumno/a es aceptable, o que decida cuál de varios posibles enunciados (conocidos como bandas de puntuación) describe mejor en qué medida se acerca a la respuesta "perfecta". En la tabla 5 se ilustra un ejemplo del tipo de bandas de puntuación que utiliza el IB.

Tabla 5
Ejemplo de bandas de puntuación del IB

Puntuación	Descriptor de nivel
0	El trabajo no alcanza ninguno de los niveles especificados por los descriptores que figuran a continuación.
1-3	<ul style="list-style-type: none"> Se demuestra una escasa comprensión de la cuestión. No se utiliza vocabulario específico de la asignatura, o se utiliza constantemente de manera inapropiada. La explicación de la cuestión es mínima. Los argumentos aportados son superficiales y con frecuencia poco claros. La respuesta es descriptiva. Cuando se incluye análisis, este es superficial o incoherente. Hay poca discusión, o ninguna en absoluto, de distintos puntos de vista. Cuando se incluye una conclusión, esta es muy superficial o no es coherente con el resto de la respuesta.
4-6	<ul style="list-style-type: none"> Se demuestra una comprensión básica de la cuestión. Se utiliza vocabulario específico de la asignatura, pero a menudo de manera inapropiada. La explicación de la cuestión es básica y está poco desarrollada. Los argumentos son a menudo imprecisos o vagos, y con frecuencia no queda claro lo que se trata de decir en la respuesta. Hay un análisis limitado y la respuesta es, en general, más descriptiva que analítica. Hay una discusión limitada de distintos puntos de vista. La conclusión incluida es simplista.
7-9	<ul style="list-style-type: none"> Se demuestra cierta comprensión de la cuestión. Se utiliza vocabulario específico de la asignatura y, a veces, de manera apropiada. Se ofrece una explicación satisfactoria de la cuestión, pero le falta claridad y desarrollo en algunas partes. Se aportan argumentos pertinentes, pero les falta precisión y detalle. La respuesta incluye análisis, pero este carece de desarrollo. Hay cierta discusión de distintos puntos de vista. Se incluye una conclusión.
10-12	<ul style="list-style-type: none"> Se demuestra una buena comprensión de la cuestión. Se utiliza vocabulario específico de la asignatura y, en general, de manera apropiada. La explicación de la cuestión es clara, aunque requiere un mayor desarrollo. Se aportan argumentos pertinentes y precisos, pero les falta detalle. La respuesta incluye un análisis crítico, pero este carece de desarrollo. Hay discusión de distintos puntos de vista. La respuesta brinda argumentos que conducen a una conclusión, y esta es coherente con los argumentos presentados.
13-15	<ul style="list-style-type: none"> Se demuestra una muy buena comprensión de la cuestión. Se utiliza vocabulario específico de la asignatura de forma correcta y precisa. La explicación de la cuestión es clara y está bien desarrollada. Se aportan argumentos pertinentes, precisos y detallados. La respuesta incluye un análisis crítico bien desarrollado. Hay discusión crítica de distintos puntos de vista. La respuesta brinda argumentos que conducen a una conclusión razonada y formulada con claridad, y esta es coherente con los argumentos presentados.

Otro aspecto que tener en cuenta es si la corrección se lleva a cabo de manera separada para distintos aspectos del trabajo, lo que a menudo se denominan **criterios**. Por ejemplo, un ensayo se puede evaluar con respecto a cuatro criterios distintos: calidad de la gramática, exactitud de los datos clave, estructura del ensayo, y calidad de la conclusión. En este enfoque, se considera una buena práctica asegurarse de que los

critérios sean independientes los unos de los otros, ya que no sería justo que un alumno/a recibiera puntos en distintas partes por el mismo aspecto de su respuesta.

Lo contrario a la **corrección por criterios** es la **corrección de impresión global**. En este caso, el examinador/a sopesa los distintos aspectos de una respuesta ideal y emite una valoración final que refleja el trabajo en conjunto.

El factor principal que puede dificultar la fiabilidad de la corrección es la naturaleza de la tarea y la amplitud de las respuestas que se esperan del alumnado. Una tarea amplia en la que se espera una gran variedad de respuestas sería más difícil de corregir de manera fiable que una tarea limitada en la que solo se recibe una escasa variedad de respuestas.

Para obtener más información sobre los diferentes enfoques de la corrección, consulte la sección ["Corrección"](#).

Corrección de la evaluación formativa

Todos los programas del IB utilizan la evaluación formativa. Sin embargo, en el Programa de la Escuela Primaria (PEP) es el **único** tipo de evaluación que se utiliza. En la evaluación formativa, la corrección no tiene por qué ser únicamente numérica. Es plenamente posible comparar un trabajo con la respuesta "perfecta" y aportar comentarios descriptivos sobre las similitudes y las diferencias entre ambos. Esto podría adoptar la forma de criterios de logro compartidos que dejen claro al alumnado qué es lo que se considera "bueno". Aunque esto puede dificultar comparar dos respuestas para determinar cuál es la mejor, el objetivo de la evaluación formativa es aportar comentarios para respaldar el aprendizaje.

En el aula, el profesorado puede utilizar la evaluación formativa para identificar las carencias en los conocimientos, la comprensión o las habilidades de un alumno/a. En tal caso, la fiabilidad es menos importante. No es necesario otorgar ninguna puntuación o calificación. El resultado de la evaluación podría ser una lista de temas o habilidades en los que el alumno/a debe trabajar. En esta situación, la pertinencia del constructo es importante porque el alumno/a necesita recibir comentarios sobre los aspectos que son importantes para el programa de estudios o para permitirle demostrar lo que sabe.

Fomento de una mentalidad ética

Definición de la mentalidad ética

La mentalidad ética es el compromiso de actuar de forma **responsable** y **auténtica**, guiándose por los principios de **honestidad**, **respeto** e **integridad**. En un contexto educativo, fomenta la integridad académica, haciendo hincapié en la importancia de generar trabajos legítimos y honestos. Esta mentalidad trasciende las normas y políticas, e integra la conducta y la toma de decisiones éticas en la cultura de una comunidad de aprendizaje. Fomenta que el alumnado valore su proceso de aprendizaje, que tome decisiones basadas en principios, y que contribuya a un entorno académico justo.

Por qué generar una mentalidad ética

Desarrollar una mentalidad ética es esencial para garantizar la credibilidad y la validez de los logros educativos. La mentalidad ética protege la integridad de las evaluaciones, lo cual permite que los logros del alumnado reflejen sus habilidades reales. Además, fomentar el comportamiento ético prepara al alumnado para la vida más allá del aula, al inculcarle valores fundamentales para tener éxito en la educación superior y en los entornos profesionales. Cuando se hace caso omiso a los principios éticos, la confianza dentro de la comunidad escolar se erosiona, y la credibilidad de la institución se pone en peligro. Al fomentar una mentalidad ética, los colegios facultan al alumnado para formar parte de una ciudadanía global como personas íntegras y comprometidas con la equidad y el respeto en todas las facetas de la vida.

Cómo crear una mentalidad ética

Desarrollar una mentalidad ética implica un enfoque holístico que abarca a toda la comunidad. Los colegios deben integrar la integridad académica en su cultura, comenzando por comunicar claramente las expectativas específicas de cada etapa de desarrollo del alumnado. El profesorado desempeña un papel fundamental al dar ejemplo de comportamiento ético, diseñar estrategias preventivas, e incorporar discusiones sobre la integridad en los currículos. El personal de coordinación y los equipos directivos de los colegios deben asegurarse de que las políticas estén bien definidas, se apliquen de manera coherente, y se respalden mediante capacitación para todas las partes interesadas. Tener una comunicación abierta con padres, madres y tutores/as legales ayuda a alinear las expectativas, y refuerza la importancia de la ética en casa.

Los talleres periódicos, los ejemplos claros de desafíos éticos, y el fomento del diálogo sobre las aplicaciones de la integridad en el mundo real pueden ayudar a reforzar estos principios. Al hacer hincapié en los beneficios de la honestidad, en lugar de centrarse únicamente en las penalizaciones, los colegios crean un entorno alentador en el que el alumnado desarrolla las habilidades y la comprensión necesarias para la toma de decisiones éticas.

Una mentalidad ética no es solo un objetivo, sino una práctica continua que se logra mediante la responsabilidad compartida, la transmisión de mensajes coherentes, y el firme compromiso de toda la comunidad escolar. Para obtener más información, consulte el recurso para imprimir [“Desarrollo de la mentalidad ética”](#).

Uso ético de la inteligencia artificial en la evaluación

Hay dos aspectos principales que hay que tener en cuenta al analizar cómo interactúa la inteligencia artificial (IA) con la evaluación:

- ¿Qué impacto tiene la IA en lo que se evalúa?

- ¿Qué papel desempeña la IA en el proceso de evaluación del alumnado?

La IA forma parte del mundo actual, y lo responsable es trabajar con ella y ayudar al alumnado del IB a utilizar esta tecnología de manera informada y ética.

“¿Qué significa decir que un trabajo es mío?”

Esta es la pregunta clave que subyace al primer desafío: ¿qué impacto tiene la IA en lo que se evalúa? Uno de los temores iniciales sobre la IA en el ámbito de la educación era que el alumnado la utilizara para “hacer trampas”, en el sentido de pedirle a la IA que le hiciera sus trabajos. También se ha afirmado varias veces que la IA puede obtener mejores resultados que el alumnado en determinadas evaluaciones. Estas inquietudes nos llevan de nuevo a las preguntas más fundamentales de la evaluación: ¿Cuál es el propósito de una pregunta específica? ¿Qué queremos que demuestre el alumnado? Datos como las fechas de un acontecimiento histórico se pueden consultar en pocos segundos en Internet. Sin embargo, no siempre resulta adecuado buscar la respuesta en Internet, por ejemplo, en una situación en la que se tiene poco tiempo, como el aterrizaje de un avión.

Lo mismo ocurre con otros usos de la IA. A veces lo que es fundamental es el contenido, y no la presentación. En este caso, puede que no sea necesario puntuar el estilo de comunicación, y la IA podría ayudar al alumnado a presentar sus ideas de forma legible. Por el contrario, al escribir un poema, la elección de cada palabra es fundamental. El alumnado no debe utilizar la IA para componer ninguna parte de un poema final, pero podría utilizarla, por ejemplo, como diccionario de sinónimos.

La pregunta educativa fundamental que plantea la prevalencia de la IA es: ¿qué debe aprender el alumnado para poder utilizarla de manera responsable y ética en su vida cotidiana? Sin duda, esta pregunta clave evolucionará con el tiempo. Mientras tanto, la respuesta implica educar a estudiantes para:

- Recibir la información con espíritu crítico y ser consciente de los posibles sesgos o conclusiones engañosas
- Hacer un seguimiento de las fuentes y referencias para asegurarse de que se utilizan de manera correcta
- Comprender qué es y qué no es su trabajo, y presentarlo con honestidad

“Hola, seré su asistente de evaluación de IA para este examen”

La IA cambiará gradualmente la forma en que se puede evaluar al alumnado. La progresión hacia la evaluación digital permite tareas mucho más interactivas, y la incorporación de la IA puede ampliar esto aún más: el alumnado podría, por ejemplo, interactuar con avatares digitales y explicarles ideas. Esta tecnología puede crear oportunidades para diferentes enfoques de la evaluación, como un mayor uso de las entrevistas, que actualmente son inviables debido a la necesidad de examinadores/as humanos con los que interactuar.

La corrección automática de respuestas objetivas (como las que se dan a preguntas de opción múltiple o de arrastrar y soltar) es una herramienta que se utiliza en la evaluación desde hace mucho tiempo. Los avances en el aprendizaje automático y la IA también amplían el potencial de la corrección automática para utilizarse en preguntas más subjetivas y respuestas complejas. Durante un futuro próximo, cualquier puntuación final de alto impacto la deberá determinar un examinador/a humano, pero la IA ofrece un gran potencial para respaldar la calidad de la corrección e identificar los trabajos que debería revisar un segundo examinador/a.

La corrección mediante IA también podría respaldar la evaluación formativa como parte de la práctica docente, y dar lugar a oportunidades de evaluación adaptativa en las que las preguntas se ajusten a la capacidad de cada estudiante en función de sus respuestas anteriores. Este enfoque se ha utilizado en el pasado para preguntas de opción múltiple, pero los avances en IA ofrecen esta posibilidad para tipos de preguntas más amplios, lo que da lugar a exámenes más personalizados.

A la hora de considerar el impacto de cualquier nueva tecnología, los factores más importantes son garantizar que el propósito de la evaluación siga siendo pertinente para los resultados del aprendizaje esperados, y que las evaluaciones en sí mismas sigan siendo válidas para dicho propósito.

Recordemos que el uso de la inteligencia artificial puede ser beneficioso para **ayudar al alumnado en su aprendizaje**, pero que no es aceptable utilizar la IA **en lugar del aprendizaje** ni entregar trabajos que no son originales.

Lecturas adicionales

- *Ideas clave para explicar el uso de las herramientas de inteligencia artificial (IA)* (hoja informativa)
- *Evaluación de 13 situaciones hipotéticas de uso de la inteligencia artificial en trabajos de clase* (versión larga)

Definición de las prácticas

- Las prácticas que se describen en esta sección engloban las evaluaciones sumativas del Programa de los Años Intermedios (PAI), el Programa del Diploma (PD) y el Programa de Orientación Profesional (POP).
- Estas no se aplican al Programa de la Escuela Primaria (PEP), en el que el Bachillerato Internacional no ofrece evaluación sumativa, ni a ninguna evaluación formativa que pueda llevar a cabo el colegio.

En esta sección se presentan las prácticas que el IB utiliza para producir los resultados del alumnado que pasa por las evaluaciones que se corrigen o se moderan externamente. No se cubre ningún otro tipo de evaluación que no realice el IB, como podría ser una evaluación que lleve a cabo un/a docente para utilizarla en el PEP.

Un **principio** establece **por qué** se hace algo, mientras que una **práctica** describe **cómo** se hace. En esta sección se explican las prácticas que el IB utiliza para asegurar la validez de los resultados de la evaluación.

Más tarde se tratan los **procedimientos** del IB que describen los pasos individuales para aplicar cada práctica. Cuando estos procesos conciernen a los colegios, se explican en detalle en la publicación *Procedimientos de evaluación* específica de cada programa.

Qué miden las evaluaciones del IB y cuál es el papel de los conocimientos previos

La intención de las evaluaciones sumativas es medir el desempeño del alumnado en lo que respecta a **conocimientos, comprensión y habilidades** a medida que completan sus estudios en el PAI, el PD y el POP. Esto tiene varias implicaciones, tales como:

- Los resultados de la evaluación sumativa reflejan el desempeño de cada estudiante en un momento dado. La evaluación sumativa no mide el potencial de un alumno/a ni la progresión de su aprendizaje.
- Las evaluaciones del IB están diseñadas para minimizar las imprecisiones que pueda provocar el bajo desempeño de un alumno/a en una pregunta o un día concretos. Para ello, generalmente se realizan múltiples exámenes a fin de brindarle varias oportunidades de demostrar lo que sabe hacer. No obstante, la cantidad de exámenes debe ser manejable y no representar una carga excesiva para el alumnado.
- La evaluación sumativa es un reflejo de los conocimientos, la comprensión y las habilidades de cada estudiante, no de un grupo de estudiantes.
- El IB no tiene en cuenta los conocimientos previos al asignar calificaciones, lo que significa que no se tienen en cuenta las titulaciones, las calificaciones ni los logros obtenidos antes de comenzar el programa pertinente. Las guías de asignaturas del IB describen la pertinencia de los conocimientos previos para una asignatura o área de aprendizaje.

Comunicación de los logros del alumnado

- La declaración de principios del IB establece el objetivo de formar personas capaces de crear un mundo mejor. Una buena evaluación deberá respaldar este objetivo.
- Las calificaciones finales del IB tienen un significado, y los límites de calificación se establecen teniendo en cuenta dicho significado.
- Si bien las calificaciones representan una visión muy simplificada de los logros del alumnado, permiten que otras partes interesadas (como las universidades, las empresas o los institutos de educación superior) dispongan de una base sobre la que realizar juicios razonables de selección.
- Si se proporcionara información más compleja y holística, entonces terceras partes deberían simplificar dicha información para tomar decisiones sensatas de selección, y es probable que no fuesen tan cuidadosas como lo es el IB al establecer los límites de calificación.
- Los logros del alumnado no se reducen a los resultados obtenidos en los exámenes, así que, incluso cuando se trate de calificaciones finales, lo que para un alumno/a puede suponer un resultado decepcionante, para otro/a podría ser un gran logro.

El Bachillerato Internacional tiene como meta formar personas solidarias, informadas y ávidas de conocimiento, capaces de contribuir a crear un mundo mejor y más pacífico, en el marco del entendimiento mutuo y el respeto intercultural.

(Declaración de principios del IB, 2024)

El significado de la calificaciones finales del IB

Los resultados de las evaluaciones del IB se denominan *calificaciones finales*. Estas calificaciones representan la calidad del trabajo que el alumno/a ha demostrado en sus respuestas.

El IB publica las descripciones de cada una de las calificaciones. Estas descripciones son distintas para el PAI, el PD y el POP, ya que reflejan la calidad del trabajo que se espera del alumnado en función de su edad.

Figura 16

Ejemplo de descriptores de calificaciones finales del PD

Descriptores de calificaciones finales del Grupo 3 (Individuos y Sociedades)

Individuos y Sociedades

Calificación 7

El alumno demuestra conciencia conceptual, perspicacia, y un conocimiento y una comprensión que resultan evidentes en las habilidades de pensamiento crítico; un nivel alto de capacidad para proporcionar respuestas que están plenamente desarrolladas, estructuradas de manera lógica y coherente, e ilustradas con ejemplos apropiados; un uso preciso de la terminología específica de la asignatura; familiaridad con la literatura de la asignatura; una capacidad para analizar y evaluar indicios y para sintetizar conocimientos y conceptos; conciencia de que existen puntos de vista alternativos y sesgos subjetivos e ideológicos, y la capacidad de llegar a conclusiones razonadas aunque inciertas; indicios sólidos de pensamiento crítico y reflexivo; un alto nivel de competencia para analizar y evaluar datos o resolver problemas.

Calificación 6

El alumno demuestra un conocimiento y una comprensión detallados; respuestas coherentes que están estructuradas de manera lógica y bien desarrolladas; un uso coherente de la terminología adecuada; una capacidad para analizar, evaluar y sintetizar conocimientos y conceptos; conocimiento de estudios de investigación, teorías y cuestiones pertinentes, y conciencia de que existen perspectivas y contextos diferentes a partir de los cuales estos fueron desarrollados; indicios sólidos de pensamiento crítico; una capacidad para analizar y evaluar datos o resolver problemas de manera competente.

Calificación 5

El alumno demuestra un buen conocimiento y una buena comprensión de la asignatura utilizando terminología específica de la asignatura; respuestas que están estructuradas de manera lógica y coherente, pero que no están plenamente desarrolladas; una capacidad para proporcionar respuestas competentes, con algún intento de integrar conocimientos y conceptos; una tendencia a ser más descriptivo que evaluativo (aunque se demuestra cierta capacidad para presentar y desarrollar puntos de vista opuestos); algunos indicios de pensamiento crítico; una capacidad para analizar y evaluar datos o resolver problemas.

Los descriptores generales de las calificaciones finales deben ser idénticos para todas las asignaturas de un grupo de asignaturas, pero a menudo el IB los coloca en contextos específicos de cada asignatura para que sea más fácil entender lo que significan en cada caso. Es importante comprender que el estándar no cambia en función del contexto de la asignatura. Una calificación de 4 en Lengua debe significar lo mismo que una calificación de 4 en Ciencias o en una asignatura de Artes. Esta idea se encuentra en el enfoque que el IB da a sus programas, donde todas las calificaciones cuentan lo mismo. Sin embargo, hay mucho debate entre la comunidad educativa sobre si este concepto tiene sentido: ¿cómo se pueden comparar los logros alcanzados en dos asignaturas distintas? ¿Tiene siquiera sentido intentarlo? Este concepto se trata con más detalle en la sección "Equiparabilidad".

Figura 17

¿Cómo pueden compararse estos dos trabajos?



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

tomando el logaritmo neperiano a ambos lados

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

(b) las coordenadas de Q son (1,0)

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

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

$$y = x - 1$$

(c) sea A el área que se pide para hallar

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

se utiliza la integración por partes $\int \ln x dx$

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

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

Como parte de las consideraciones sobre la validez, el argumento a favor de la equiparabilidad entre las calificaciones del IB en diferentes asignaturas se basa en el propósito de las calificaciones finales del IB. Las calificaciones finales del IB están pensadas para permitir que las partes implicadas puedan comparar los logros del alumnado. Por tanto, tiene sentido utilizar métodos estadísticos y cualitativos para intentar lograr la paridad en el significado de las calificaciones.

Diferencia entre puntuaciones y calificaciones

Puntuación y calificación son conceptos distintos.

Figura 18

Metáfora para comparar puntuaciones y calificaciones

Se esperaría que los "buenos" alumnos/as completaran la mayor parte de la tarea.



Se esperaría que los "buenos" alumnos/as completaran solo parte de la tarea.

Se pueden utilizar numerosas metáforas para explicar la diferencia entre puntuaciones y calificaciones. Por ejemplo, en la figura 18, la distancia recorrida podría representar la puntuación, como medida común de la distancia que ha recorrido una persona. Sin embargo, para comprender el logro que supone recorrer una distancia determinada, es necesario tener en cuenta dónde caminaba la persona: esto es lo que se tiene en cuenta al establecer las calificaciones.

- Las **puntuaciones** representan **en qué medida** el alumno/a ha completado la tarea.
- Las **calificaciones** tienen en cuenta **la dificultad** de la tarea, con el fin de cuantificar el nivel de logro que representa la puntuación del alumno/a.

Prestemos atención a los dos ejemplos que se dan en la figura 19. En el primer ejemplo se esperaría que un alumno/a de 16 años realizase casi toda la tarea correctamente para recibir una "buena" calificación. En el segundo ejemplo se esperaría bastante menos para que el alumno/a lograra una "buena" calificación.

Figura 19

Tareas de evaluación con diferentes niveles de dificultad

The image shows two musical scores side-by-side. On the left is the score for 'Three Blind Mice', featuring a simple melody with lyrics: 'Three blind mice, Three blind mice, See how they run, See how they run. They all run af-ter the fir-mer's will, who cut off their tails with a carving knife, Did you e- ver see such a sight in your life, as three blind mice?'. On the right is the score for 'SONATA XIV' by Beethoven, marked 'Allegro (♩ = 94)'. It is a complex piano sonata with multiple staves and intricate musical notation.

Se esperaría que los “buenos” alumnos/as completaran la mayor parte de la tarea.

Se esperaría que los “buenos” alumnos/as completaran solo parte de la tarea.

Esta consideración conduce a uno de los desafíos que se plantean al establecer evaluaciones de buena calidad. El alumnado debe tener la oportunidad de demostrar su pleno potencial, algo que no resulta posible en tareas demasiado simples. Por el contrario, si las tareas son demasiado difíciles, puede suceder que gran parte del alumnado no pueda ni siquiera empezar la tarea. Esto dificulta diferenciar entre el alumnado mediante las calificaciones.

Las tareas demasiado sencillas también pueden llevar a medir la precisión en vez de la comprensión. De esa manera, el alumnado que cometiera errores menores no recibiría las calificaciones más altas aunque tuviera una buena comprensión del tema.

Para obtener más información sobre la diferencia entre puntuaciones y calificaciones, consulte el video “La diferencia entre las puntuaciones y las calificaciones: cómo calibramos el grado de dificultad de una evaluación y no solo lo que los alumnos hicieron bien”.

Una convocatoria de exámenes satisfactoria

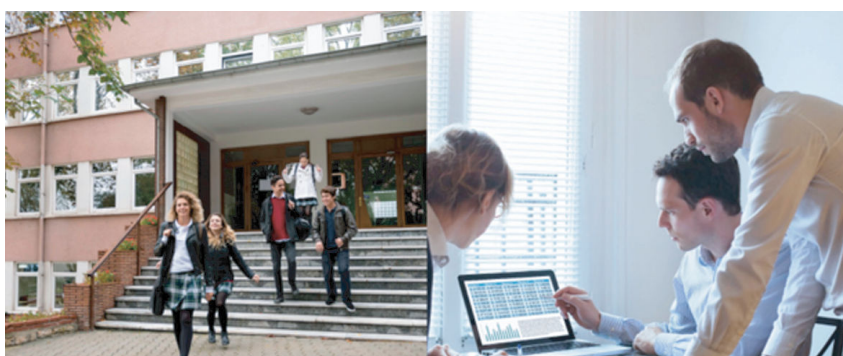
Figura 20

Diferentes perspectivas acerca de qué es una convocatoria de exámenes satisfactoria



Se esperaba que los “buenos” alumnos/as completaran solo parte de la tarea.

Pudimos elegir al alumnado adecuado para nuestros cursos.



A nuestro alumnado le ha ido mejor que al de los colegios cercanos.

No hubo errores en las preguntas ni en la corrección.

Determinar si algo es satisfactorio depende del punto de vista de cada persona. El IB hace hincapié tanto en la pertinencia general del constructo como en la facilidad de uso: Para que se considere satisfactoria, una convocatoria de exámenes debe cumplir los siguientes requisitos:

- La evaluación debe dar a todo el alumnado una oportunidad justa para demostrar sus capacidades.
- La experiencia debe ser lo más sencilla posible para los colegios y para el alumnado.
- Todas las partes interesadas deben confiar en los resultados (las calificaciones finales) que el IB publica.

Calificaciones y logros

Como se refleja en la declaración de principios, el objetivo de la educación del IB va mucho más allá de una serie de calificaciones académicas. Esto queda reflejado en el perfil de la comunidad de aprendizaje del IB y se explica en la publicación *¿Qué es la educación del IB?* (2019).

Ahora bien, los resultados de la evaluación solamente pueden centrarse en una pequeña parte de los aspectos de la declaración de principios. Teniendo en cuenta el modelo de compensación utilizado en los exámenes del IB, ¿qué proporción de los puntos que se conceden en la evaluación de Matemáticas se debería atribuir a la cualidad de *solidaridad* o de *audacia* (suponiendo que sea razonable darles un valor

numérico a estas cualidades esenciales)? Uno de los principios de la evaluación del IB es que debe producir un efecto de repercusión positivo que facilite alcanzar los objetivos del IB.

Figura 21

Visión holística de los logros del alumnado



Al comunicar los resultados de un programa del IB, es importante no tener en cuenta solamente las calificaciones finales para ilustrar la totalidad de los logros del alumno/a.

En las evaluaciones, el IB solo registra los logros finales de cada estudiante, sin precisar lo difícil que le resultó alcanzar ese resultado, o su pleno potencial. El IB es consciente de la importancia de estos dos elementos, pero considera que no es posible medirlos correctamente en la evaluación sumativa. La responsabilidad de realizar esta evaluación recae en el colegio, que es quien tiene una visión holística de cada estudiante.

Si bien existen métodos que permiten calcular el “valor añadido” de los logros del alumnado o “calificaciones previstas basadas en los logros anteriores”, el IB recomienda proceder con cautela cuando se utilicen para hacerse una idea del éxito de cada estudiante. Estas mediciones se basan en un alumno/a promedio, pero cabe recordar que cada estudiante representa una combinación única de características y rasgos personales que deberían reconocer y valorar de manera adecuada aquellas personas que tienen la oportunidad de aprender con el alumno/a durante un programa completo.

Calificaciones previstas

Como se indica en la publicación *Normas para la implementación de los programas y aplicaciones concretas* (2020), “El colegio comunica las calificaciones previstas del IB [...] al IB como parte de la evaluación validada por el IB, estipuladas de modo que valoren la integridad académica y reflejen la máxima precisión posible”.

El desempeño del alumnado puede variar durante su trayectoria educativa, lo que dificulta prever sus calificaciones finales. Las calificaciones previstas cumplen varias funciones importantes. Aportan pruebas durante las reuniones de concesión de calificaciones, lo cual ayuda al IB a determinar el nivel general de fortaleza del grupo de estudiantes y la distribución de las calificaciones para cada asignatura. También ayudan a identificar grandes discrepancias entre las calificaciones previstas y las reales, lo cual permite realizar controles de calidad adicionales si resulta necesario. Para los colegios, las calificaciones previstas ayudan a establecer y gestionar las expectativas del alumnado y de sus padres, madres y tutores/as legales.

Aunque es poco probable que un profesor/a pueda determinar con total precisión todas las calificaciones previstas de sus estudiantes, el objetivo será alcanzar la mayor precisión posible. Cuanto más precisas sean las calificaciones previstas, más útiles y constructivas resultarán dentro del proceso.

Una calificación prevista es un punto de datos importante para la evaluación: se tiene en cuenta a nivel general de la asignatura para establecer los límites de calificación, y también se tiene en cuenta a nivel escolar, siempre que haya pruebas de su precisión. Por tanto, es importante que las calificaciones previstas sean precisas para que puedan utilizarse como puntos de datos fiables.

Una calificación prevista también es:

- Un punto de datos para los colegios, que pueden elegir si compartirlas con el alumnado o no
- Una indicación de capacidad para las opciones posteriores al programa, con el fin de preparar al alumnado para tener éxito
- Una calificación (1-7, o A-E), no una puntuación

A la hora de predecir calificaciones, es importante comprender qué no es una calificación prevista. Una calificación prevista no es:

- Un mecanismo para compensar circunstancias adversas
- Una herramienta disciplinaria
- Una ecuación basada en los límites de calificación de una convocatoria anterior
- Equivalente a otros usos de las calificaciones previstas, como el acceso a la universidad o a otras instituciones de educación superior
- Un mecanismo para tranquilizar al alumnado y a sus padres, madres o tutores/as legales
- Una exageración de las capacidades del alumnado en cualquier dirección

Para obtener más información sobre las calificaciones previstas y sobre estrategias para mejorar la precisión, consulte el recurso para imprimir "[Calificaciones previstas: guía para el profesorado](#)" y la siguiente orientación disponible en el Centro de Recursos para los Programas:

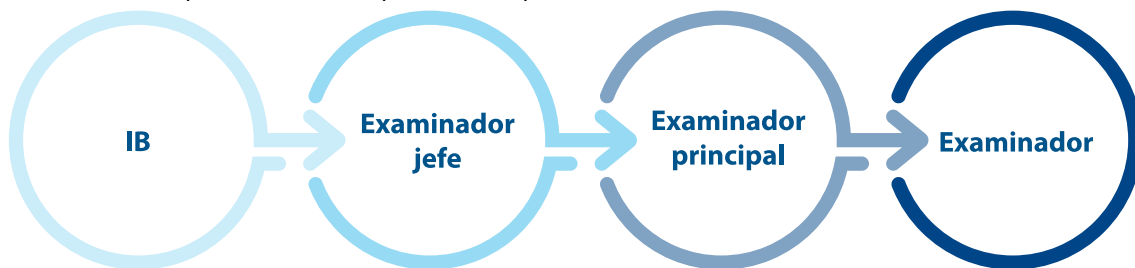
- **Recursos del PAI > La evaluación electrónica del PAI > Recursos para todas las convocatorias > Calificaciones previstas del IB**
- **Recursos del PD > Evaluación > Recursos para todas las convocatorias > Calificaciones previstas del IB**
- **Recursos del POP > Evaluación > Recursos para todas las convocatorias > Calificaciones previstas del IB**

El proceso de evaluación: funciones y responsabilidades

- Cada una de las funciones del proceso de evaluación conlleva sus propias responsabilidades y habilidades.
- En algunos casos, es posible que una misma persona asuma distintas funciones en distintos momentos del ciclo de evaluación.
- El IB lleva a cabo algunas de estas funciones, pero otras se confían a personas expertas de la comunidad del IB. En este último caso, el IB es el responsable de la decisión final y de mantener la calidad de todos los aspectos de las evaluaciones.

Figura 22

Responsabilidades de quienes desempeñan funciones clave en el ciclo de evaluación



IB

- Responsable de todos los aspectos de la evaluación
- Responsable de los procesos de evaluación, entre los que se incluyen la contratación de personal examinador, el control de calidad del trabajo del personal examinador, y la publicación de resultados
- Toma decisiones sobre cuestiones relacionadas con conductas improcedentes del alumnado, mala administración por parte de los colegios, adecuaciones inclusivas de acceso, y circunstancias adversas
- Acepta o rechaza los límites de calificación que recomienda el examinador/a jefe

Examinador principal

- Responsable de un componente
- Toma la decisión final sobre qué puntuación deben recibir las respuestas del alumnado en ese componente
- Se asegura de que todos los examinadores/as comprendan el estándar de corrección en ese componente
- Guía al examinador/a jefe para establecer límites de calificación para el componente

Examinador jefe

- Tiene una visión de conjunto de todos los componentes de su asignatura o grupo de asignaturas
- Garantiza la coherencia de los estándares entre todos los componentes, incluida la redacción de las pruebas de examen
- Media en los problemas académicos relacionados con la evaluación
- Recomienda al IB los límites de calificación finales

Examinador

- Corrige el trabajo del alumnado siguiendo el estándar que fija el examinador/a principal

Examinador principal y examinador jefe

El examinador/a principal:

- Gestiona un componente específico, como la evaluación interna
- Establece los estándares de corrección en una convocatoria y entre convocatorias, y los explica al equipo examinador mediante discusiones de estandarización y modelos de calidad
- Orienta al examinador/a jefe con recomendaciones sobre los límites de calificación para el componente del que es responsable
- Suele crear contenidos de evaluación, a menos que exista un conflicto de intereses
- Trabaja para el IB durante la convocatoria de exámenes como profesional educativo externo, y suele contribuir en varias convocatorias

El examinador/a jefe es responsable de mantener la calidad de varios componentes relacionados entre sí. En el PD y el POP, es responsable de una asignatura completa, mientras que en el PAI (que normalmente solo tiene un componente por disciplina), el examinador/a jefe es responsable de un grupo de asignaturas, como Ciencias o Adquisición de Lenguas. El examinador/a jefe actúa como especialista académico del IB en su área.

El examinador/a jefe:

- Garantiza la calidad de los componentes relacionados con una asignatura completa (PD y POP) o un grupo de asignaturas (PAI)
- Actúa como especialista académico del IB, mantiene los estándares y garantiza la coherencia
- Ayuda a resolver desacuerdos
- Dirige el proceso de concesión de calificaciones y recomienda al IB los límites de calificación
- A menudo también actúa como examinador/a principal de un componente determinado
- Trabaja para el IB durante un período específico

En el caso de las asignaturas con poco alumnado, el IB designa examinadores/as para que desempeñen la función de examinador/a responsable. Los examinadores/as responsables tienen las mismas responsabilidades que los examinadores/as jefe en relación con su asignatura, que suele contar con menos examinadores/as principales y personal examinador. Además, se invita a los examinadores/as jefe a colaborar con el IB en la discusión y la mejora de los currículos y las evaluaciones mediante actividades de revisión del currículo, que no se tratan en esta publicación.

Otros examinadores

El personal examinador es responsable de corregir el trabajo entregado para evaluación según el estándar que fija el examinador/a principal. Para garantizar que comprenden y aplican dicho estándar, se utiliza el modelo de calidad.

Los examinadores/as interesados solicitan trabajar con el IB para realizar la corrección en una convocatoria de exámenes. Dichos examinadores/as deben ser personas expertas en la asignatura que quieran corregir (normalmente serán docentes de esa asignatura). Además, deben tener experiencia en la enseñanza a estudiantes de la franja de edad correspondiente. El IB verifica las credenciales mediante un riguroso proceso de solicitud que incluye la presentación de referencias. Por lo general, al personal examinador solamente se le ofrece corregir un componente de trabajo de clase y un componente de examen por convocatoria. El período de corrección de esos dos componentes no se solapa. Al personal examinador se le paga por cada examen con las respuestas del alumno/a en condiciones reales. A efectos de la remuneración, las respuestas de aptitud y las de control (véase la sección “El modelo de calidad”) no se tienen en cuenta, porque su función es demostrar al IB que los examinadores/as comprenden el estándar de corrección.

Los jefes/as de equipo son examinadores/as con mucha experiencia a quienes el IB pide que ayuden a otros examinadores/as a comprender el estándar de corrección. Los jefes/as de equipo prestan apoyo al personal examinador durante el proceso de corrección de respuestas de aptitud, y ofrecen comentarios si las respuestas de control indican que se están desviando del estándar definitivo. El número de jefes/as de

equipo en una asignatura dependerá del número total de examinadores/as, que a su vez depende del tamaño del grupo de estudiantes. En una asignatura con poco alumnado, el examinador/a principal puede prestar apoyo a todo el personal examinador.

El equipo de estandarización suele estar compuesto por examinadores/as con experiencia que participan en las discusiones cualitativas que conducen al desarrollo de un modelo de calidad. El equipo de estandarización analiza la aplicación del estándar del examinador/a principal para garantizar que se pueda aplicar y cumplir.

Para obtener más información sobre el rol de los examinadores/as, consulte el video [“Presentación de los examinadores: quiénes son y qué hacen”](#).

Las responsabilidades del personal del IB

Responsables de asignaturas y analistas de operaciones de evaluación son algunas de las personas que participan en garantizar una evaluación fiable desde la perspectiva del IB.

En el IB, varios equipos gestionan las solicitudes de modificación de pruebas de examen y otras adecuaciones inclusivas de acceso, así como las solicitudes por circunstancias adversas, las cuestiones relacionadas con la integridad académica, y la supervisión de la calidad del trabajo del personal examinador y de los factores de moderación de los colegios.

El director/a de evaluación, con el apoyo del jefe/a de desarrollo y logística de la evaluación y del jefe/a de principios y prácticas de evaluación, es la persona responsable de estudiar los límites de calificación que recomienda el examinador/a jefe como culminación de una reunión de concesión de calificaciones para cada asignatura, y de decidir si está conforme o pedirle al examinador/a jefe que revise la recomendación.

En última instancia, el IB es responsable de todas las decisiones que se toman como parte del ciclo de evaluación. Aunque emplee personal experto externo para elaborar evaluaciones justas y de calidad, la responsabilidad final recae sobre el IB.

Funciones relacionadas con la creación de exámenes

Los creadores/as de contenidos de evaluación tienen una función específica dentro del proceso de creación de exámenes, que puede incluir el trabajo de autor/a, revisor/a o verificador/a. A continuación se indican las habilidades clave que deben tener los creadores/as de contenidos de evaluación:

- Creatividad, para elaborar preguntas interesantes y que se ajusten al currículo
- Conciencia cultural, para asegurarse de que las preguntas no contengan sesgos
- Habilidades de comunicación, para escribir preguntas claras y sin ambigüedades que se traduzcan bien y que evalúen las habilidades pertinentes, de manera que conduzcan a correcciones fiables

Otras personas implicadas en la elaboración de evaluaciones son, por ejemplo:

- Diseñadores/as técnicos, que dan formato a las preguntas para los exámenes
- Traductores/as, que convierten las evaluaciones a las lenguas requeridas
- Revisores/as externos, que realizan la evaluación como si fueran estudiantes para proporcionar comentarios sobre la extensión y la claridad de esta

Para obtener más información sobre cómo se crean las evaluaciones, consulte la sección [“Elaboración de evaluaciones”](#).

Jerarquía del personal examinador

- El examinador/a principal es quien tiene la decisión final sobre la puntuación que se concede a un componente. Su estándar se considera el definitivo. El resto del personal examinador tiene que seguir el estándar definitivo.
- El IB parte de la base de que las personas que han colaborado para establecer el estándar del examinador/a principal comprenden mejor dicho estándar que quienes lo han aprendido mediante el modelo de calidad y los jefes/as de equipo.

- El adjetivo *supervisor* implica proximidad al examinador/a principal en la jerarquía del personal examinador, y no necesariamente una larga experiencia como examinador/a o como docente.

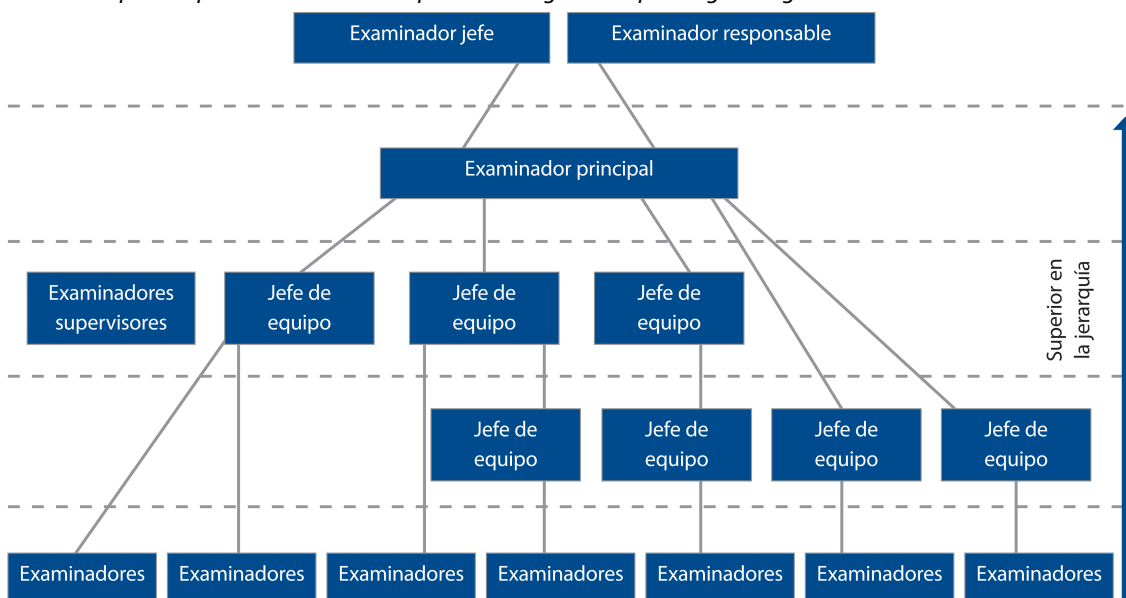
El principio por el que se rigen las correcciones en el IB es que el examinador/a principal tiene la última palabra sobre cuál es la puntuación correcta que se debe conceder al trabajo de un alumno/a. Aunque hay que reconocer que puede haber definiciones muy diferentes e igualmente válidas de qué es una puntuación “correcta”, no es justo para el alumnado obtener puntuaciones distintas en función de quién sea el examinador/a que corrija su trabajo. Por eso, el examinador/a principal establece el estándar y pide a los demás examinadores/as que lo sigan.

El examinador/a principal cuenta con el apoyo de un equipo examinador supervisor para fijar el estándar. Como estos examinadores/as tienen la oportunidad de discutir el estándar con el examinador/a principal, el IB considera que serán quienes mejor lo entiendan. Sin embargo, para garantizar la fiabilidad de la corrección, en cada convocatoria de exámenes se analiza en qué medida el personal examinador ha corregido de acuerdo con el estándar del examinador/a principal.

Existe una jerarquía del personal examinador. Para una asignatura que tenga una gran cantidad de estudiantes, la estructura de la jerarquía podría ser similar a la que se presenta en la figura 23.

Figura 23

Jerarquía del personal examinador para una asignatura que tenga una gran cantidad de estudiantes



La jerarquía puede ser flexible. Si los datos procedentes de la corrección de respuestas de aptitud o de control muestran que un examinador/a comprende el estándar del examinador/a principal mejor que un miembro con más nivel en la jerarquía, esto se tendrá en cuenta. Un jefe/a de equipo no solo debe saber aplicar el estándar, sino también ser capaz de explicarlo. Por tanto, la fiabilidad no es la única consideración a la hora de designar a los jefes/as de equipo.

En una situación ideal, no haría falta tomar decisiones basándose en la jerarquía, porque todo el equipo examinador corregiría siguiendo el mismo estándar. Sin embargo, en la realidad este es un método fundamental para determinar qué puntuaciones debe utilizar el IB. Si se da un caso especialmente difícil o controvertido, el examen se remite al examinador/a principal para que determine la puntuación adecuada.

Integridad de la evaluación

- Las evaluaciones del IB solamente pueden ser justas si se da igualdad de oportunidades de logro a todo el alumnado.
- La mala administración por parte de los colegios y las acciones de estudiantes que contravienen la política de integridad académica ponen en situación de desventaja al alumnado que ha seguido las normas. El IB hace todo lo posible para evitar tales comportamientos.
- Las normas del IB, que se detallan en la sección “Reglamento general” de los *Procedimientos de evaluación* de cada programa y en la *Política de integridad académica*, están diseñadas para minimizar las posibilidades de que haya conducta impropia del alumnado o mala administración por parte de los colegios. En última instancia, el colegio es el único que puede crear una cultura de aprendizaje en la que la conducta impropia se rechace y se denuncie.
- Si bien ciertas formas de evaluación son menos vulnerables a la conducta impropia que otras, el IB mantiene como principio que la pertinencia del constructo (es decir, evaluar lo que realmente se quiere evaluar) debe ser la prioridad principal al diseñar las evaluaciones.
- Se debe informar al IB sobre cualquier sospecha de conducta impropia del alumnado o de mala administración por parte del colegio

Definición de la integridad académica

La integridad académica es un principio rector en el ámbito educativo y consiste en tomar la determinación, como individuo, de comportarse de una manera responsable y que inspire la confianza de los/as demás. La integridad académica es la base de la conducta y la toma de decisiones éticas en el aprendizaje, la enseñanza y la evaluación, así como en la creación de trabajos académicos legítimos, de autoría original y honestos. No se debe considerar como una serie de normas estrictas que se tienen que imponer, sino que debe adoptar la forma de una cultura positiva en el seno del colegio y en la comunidad más amplia (incluidos padres, madres y tutores/as legales), que conduzca de manera orgánica a resultados de evaluación más justos para todos.

Para que una evaluación sea válida, debe reflejar de manera precisa los logros de un alumno/a en comparación con el resto del alumnado que ha realizado la evaluación. Por ese motivo, el IB vela por ofrecer enfoques coherentes para la corrección, la calificación y la eliminación de sesgos en sus exámenes. Las normas y reglamentos que establece el IB son otro aspecto para crear igualdad de oportunidades y garantizar la equidad.

En la *Política de integridad académica* se define la conducta impropia como “toda acción (ya sea deliberada o involuntaria) de un alumno/a por la cual este o un tercero sale o puede salir beneficiado injustamente en uno o varios componentes de la evaluación”. Tal actividad no afecta solamente al alumnado que se ve implicado, sino a todas las personas que han realizado la evaluación al mismo tiempo que este, ya que reduce la validez de la evaluación. Por ese motivo, el IB se toma muy en serio la conducta impropia. En la *Política de integridad académica* se encuentra información más detallada sobre cómo prevenir la conducta impropia y sobre sus consecuencias.

El diseño de la evaluación es una herramienta importante para evitar la conducta impropia, porque algunos tipos de evaluación resultan más fáciles de controlar que otros. Por ejemplo, en un examen es más difícil obtener ayuda de otra persona que en un trabajo de evaluación interna. Si bien el IB tiene en cuenta estos elementos al definir el enfoque de la evaluación, se parte del principio de que la pertinencia del constructo no se debe sacrificar para evitar las conductas impropias.

La manejabilidad es otro aspecto importante de la validez, y se debe tener en consideración al tratar de impedir que haya conductas impropiedades. La preparación de las salas de examen, en particular en el caso de la evaluación digital, puede suponer un auténtico desafío a los colegios, por lo que es un aspecto que no se debe descuidar. Al establecer las distintas normas, el IB procura tener en mente la experiencia de los colegios, y agradece que el personal de coordinación y el equipo directivo del colegio le comuniquen qué prácticas han encontrado óptimas y cuáles les han resultado complicadas.

Para obtener más información sobre la integridad académica y qué significa comportarse de manera ética, consulte el video ["Tomar las decisiones correctas sobre la integridad académica en el IB"](#).

Conflictos de intereses

En el IB se controla minuciosamente el acceso al contenido de examen, y la organización gestiona de manera activa cualquier conexión con estudiantes que vayan a presentarse a los exámenes que pudiera suponer un conflicto de interés. En el caso de que ocurra un conflicto de este tipo, o pudiera parecer que lo hay, se reorganizan las responsabilidades del personal.

El IB parte del principio de que ningún examinador/a puede corregir el trabajo de su propio alumnado, ni tampoco el del alumnado de colegios con los que tenga algún conflicto de interés (por ejemplo, un colegio en el que imparta otra asignatura, o un colegio del que se haya marchado recientemente). En los casos extremadamente raros en los que esto fuese inevitable por circunstancias que escaparan del control del IB (por ejemplo, que no hubiera nadie más cualificado para corregir siguiendo el estándar requerido), un segundo examinador/a independiente revisaría la corrección. El IB debe poder determinar que nada indica que se haya aplicado un estándar distinto al propio alumnado del examinador/a.

Gestión de la mala administración por parte de los colegios y de las conductas impropiedades del alumnado

Para cada programa, en la sección "Reglamento general" de los *Procedimientos de evaluación* y en las publicaciones *Instrucciones sobre la realización de los exámenes* y *La realización de los exámenes en pantalla del Programa de los Años Intermedios* se establecen las normas e instrucciones para minimizar el riesgo de mala administración. En el caso de los exámenes en papel, la publicación *Almacenamiento seguro de materiales de examen confidenciales del IB* proporciona instrucciones sobre el almacenamiento de los materiales confidenciales de la organización dentro de los colegios y las medidas que estos deben adoptar para implementar dicha política.

Los exámenes deben supervisarse de acuerdo con las instrucciones proporcionadas en la publicación *Instrucciones sobre la realización de los exámenes* (PD y POP) y *La realización de los exámenes en pantalla del Programa de los Años Intermedios* (PAI). Los supervisores/as de exámenes deben vigilar durante todo el examen para asegurarse de que no se produzca ninguna conducta impropiedad.

El IB realiza visitas sin previo aviso a los exámenes durante las convocatorias para comprobar que las prácticas obligatorias con respecto a los exámenes se llevan a cabo. Estas inspecciones son controles aleatorios de la implementación de los procesos del IB por parte del colegio. No obstante, el IB asigna una gran responsabilidad al equipo directivo y al personal de coordinación de los programas para prevenir casos de mala administración, ya que son quienes están en mejor posición para garantizar el cumplimiento diario de los estándares de calidad en sus colegios. Por otro lado, deben encargarse de que la cultura escolar de su colegio aliente a sus docentes y a sus estudiantes a adoptar las mejores prácticas y la mayor integridad posibles.

Además de mantener los estándares de integridad académica durante los exámenes, es importante garantizar que el trabajo que el alumnado realiza en clase (y en casa) sea obra suya y que no se produzcan casos de plagio.

Existen numerosos sitios web y herramientas de inteligencia artificial (IA) que ofrecen "ayuda" al alumnado con su trabajo. La mejor defensa contra este tipo de conducta impropiedad es contar con el profesor/a,

que es quien ha trabajado con el alumno/a y puede identificar si un trabajo entregado no se corresponde con el nivel habitual del alumno/a. Por este motivo, se pide tanto al alumno/a como al profesor/a que confirmen que el trabajo es obra del alumno/a. No basta con utilizar software de control de plagio: el profesorado también deben trabajar con sus estudiantes cuando redacten sus trabajos de evaluación interna, con el fin de comprobar la autenticidad de dichos trabajos. Una vez que se ha entregado el trabajo, el IB no está obligado a aceptar otras versiones si se descubre que se ha producido un plagio.

No se acepta ningún nivel de plagio: todas las citas utilizadas y el uso de herramientas de IA deben documentarse adecuadamente, tal como se establece en las publicaciones *Uso eficaz de citas y referencias* y *Política de integridad académica*. El IB utiliza varias herramientas informáticas para identificar posibles casos de conducta impropia. Cuando se encuentran pruebas, se inicia una investigación formal.

Todos los Colegios del Mundo del IB deben cumplir el requisito de establecer y gestionar una cultura de integridad académica; si se producen infracciones repetidas en este sentido, el colegio se enfrenta a consecuencias en lo que respecta a la autorización para impartir programas del IB.

Dificultades provocadas por los exámenes internacionales y las zonas horarias

El IB tiene colegios en todo el mundo, lo cual implica algunas dificultades específicas. Al contrario de lo que ocurre en la mayoría de los sistemas nacionales, no todo el alumnado del IB realiza los exámenes al mismo tiempo. Esto implica que parte del alumnado habrá acabado sus exámenes antes de que otra parte los haya empezado. Por ese motivo, el IB requiere a estudiantes y a docentes integridad para no beneficiarse de esta situación. En los casos en que hay suficientes estudiantes, se crean distintas zonas de examen con el fin de reducir el tiempo entre el fin de un examen para el alumnado de una zona horaria y el inicio del mismo examen para el alumnado de otra. Sin embargo, puede haber una considerable variación en la hora de comienzo de un examen.

El IB ha adoptado varias medidas para mitigar este riesgo, entre ellas la supervisión cuidadosa de los sitios web y las plataformas de redes sociales en los que se podría compartir el contenido de los exámenes. La regla que prohíbe sacar los cuestionarios de examen de la sala pretende precisamente limitar las posibilidades de compartir en línea las preguntas. Las normas de supervisión que estipulan que se debe supervisar al alumnado durante al menos dos horas desde el inicio de un examen por la mañana o por la tarde —si el examen dura menos de dos horas y es el único examen de esa mañana o esa tarde— también contribuyen a reducir las oportunidades de compartir información de manera prohibida.

El IB también se asegura de que las preguntas de los exámenes estén diseñadas para evaluar la comprensión más que la capacidad de recordar conocimientos, lo que implica que conocer las preguntas de antemano tendría un valor limitado. La última medida de mitigación es prohibir al alumnado hablar sobre el examen durante 24 horas después de realizarse, lo cual suprime la posibilidad de “charlas inocentes” al respecto.

El IB continúa esforzándose por innovar en esta área con el fin de reducir el riesgo de que se den conductas ilícitas debidas a las distintas zonas horarias.

Beneficios de la evaluación digital

La evaluación digital ofrece muchos beneficios para la gestión de conductas impropias. La cantidad de tiempo en que se puede acceder al contenido del examen está muy limitada, y el IB puede controlar cuándo se accede al contenido del examen y quién lo hace.

Además, cualquier modificación al examen que realice el colegio —como añadir tiempo o pausas— queda registrada y, cuando corresponde, debe justificarse. Esto significa que el IB puede delegar más responsabilidad en los colegios para que puedan realizar modificaciones razonables y al mismo tiempo seguir garantizando prácticas justas.

La evaluación digital también permite comprender con mayor detalle la experiencia de cada estudiante en el examen mediante el uso de metadatos, que pueden utilizarse como prueba para respaldar una

investigación sobre un presunto incidente de conducta impropia. Además, permite comprobar a gran escala la existencia de respuestas sospechosamente similares, algo que no es posible con los exámenes manuscritos.

El IB es consciente de que la evaluación digital se puede enfrentar a nuevas formas de conducta impropia, en particular a intentos de piratería informática, y dispone de diversas herramientas y procesos para contrarrestarlas.

Recursos

Docentes, colegios y estudiantes son quienes se encuentran en la mejor posición para prevenir y detener los casos de conducta impropia, ya que pueden crear una cultura en la que dichas prácticas sean inaceptables, y también prestar atención para abordarlas cuando ocurran. Este tema se analiza con más detalle en la sección [“Fomento de la mentalidad ética”](#) de esta publicación. Los materiales para ayudar a los colegios a crear una cultura de integridad académica (como la publicación *Política de integridad académica*), están disponibles en el sitio web del IB, en la sección [“Integridad académica”](#).

Si desea recibir más apoyo, si sospecha de algún caso de conducta impropia que no se esté abordando en su colegio, o si tiene alguna idea sobre cuestiones nuevas y emergentes relacionadas con la integridad académica, comuníquese con [“El IB Responde”](#).

Priorización de la equidad para todo el alumnado

- Para mantener su validez, la evaluación formativa y la evaluación sumativa deben tener en cuenta la variabilidad del alumnado y no discriminar a ningún alumno/a. Las evaluaciones del IB se diseñan teniendo en cuenta esta variabilidad desde el principio.
- La mejor manera de garantizar la equidad es diseñar evaluaciones que sean accesibles para todos. Las evaluaciones del IB siguen los principios del diseño universal, con lo que eliminan las barreras que podrían afectar a su validez. Los colegios deben seguir las directrices que establece el IB para garantizar que las condiciones en las que se realizan los exámenes escritos, orales y digitales no supongan una ventaja o desventaja para ningún alumno/a.
- En la *Política de acceso e inclusión* se abordan situaciones en las que se necesitan adecuaciones o medidas adicionales para facilitar el acceso a la evaluación, como son las pruebas de examen modificadas.
- En caso de circunstancias imprevistas que puedan afectar al acceso o a la validez de la evaluación, en la *Política de circunstancias adversas* se recoge una serie de medidas para mitigar los efectos negativos de tales acontecimientos.
- Tanto la fase de aprendizaje y enseñanza como la evaluación formativa deben estar en consonancia, con el fin de preparar al alumnado para las evaluaciones sumativas o del IB. Dado que uno de los principios clave al proporcionar adecuaciones inclusivas de acceso es que deben utilizarse durante el aprendizaje, la enseñanza y la evaluación (como forma habitual de trabajo del alumno/a), esta consonancia cobra una especial importancia. La consonancia también es esencial en todos los medios de evaluación, incluida la evaluación digital.
- El propósito de estas adecuaciones es garantizar la equidad para todo el alumnado. A la hora de tomar decisiones, el IB debe considerar lo que es justo para todos.

Todo el alumnado debe tener la posibilidad de demostrar sus capacidades en las condiciones de evaluación más equitativas que sean posibles. Algunas condiciones normales de evaluación pueden representar una desventaja para algunos alumnos/as, al impedirles demostrar su nivel de conocimiento y capacidad. Los acontecimientos adversos o las circunstancias ajenas a su control también pueden afectar su desempeño. Ambos factores deben tenerse en cuenta para aplicar medidas que garanticen la equidad.

La mejor forma de garantizar la equidad en una evaluación es que todo el alumnado realice la misma evaluación en las mismas condiciones. Para lograrlo, las evaluaciones se elaboran analizando detenidamente las diversas necesidades que pueden tener. El enfoque más eficaz para abordar la variabilidad del alumnado y cubrir sus necesidades es diseñar evaluaciones basadas en los principios del diseño universal para el aprendizaje y del diseño universal de evaluación. Una vez que las evaluaciones se elaboran siguiendo estos principios, el siguiente paso es garantizar la accesibilidad mediante adecuaciones inclusivas de acceso, tal y como se indica en la *Política de acceso e inclusión* del IB.

Las adecuaciones inclusivas de acceso están diseñadas para eliminar barreras y proporcionar acceso a todo el alumnado. Dichas barreras pueden surgir cuando la evaluación no tiene en cuenta o no se adapta adecuadamente a la neurodiversidad, las discapacidades, los desafíos, o las diversas circunstancias socioemocionales.

Se considerará la realización de ajustes razonables para ciertos alumnos/as con el fin de satisfacer sus necesidades específicas. Para obtener más información, consulte el documento *Política de acceso e inclusión*.

Se consideran circunstancias adversas aquellas situaciones ajenas a la voluntad del alumno/a que podrían tener un efecto negativo en su desempeño. En la *Política de circunstancias adversas* se describe el proceso para abordar y mitigar dichos efectos. Las adecuaciones inclusivas de acceso se establecen para eliminar las

barreras y garantizar que todo el alumnado tenga acceso a la evaluación. Estas adecuaciones no proporcionan una ventaja a los alumnos/as que se enfrentan a tales barreras, sino que posibilitan un acceso equitativo, de modo que permiten a todos demostrar sus habilidades, conocimientos y capacidades en las mismas condiciones.

Principios para las adecuaciones inclusivas de acceso

Los principios para las adecuaciones inclusivas de acceso se establecen en la *Política de acceso e inclusión*. La evaluación debe tener en cuenta las etapas de desarrollo de las personas jóvenes, que se reflejan en los objetivos generales de cada programa. Las adecuaciones inclusivas de acceso deben tenerse en consideración en todos los programas, tanto en la evaluación formativa como en la sumativa, y a lo largo de toda la experiencia de aprendizaje y enseñanza.

El IB debe garantizar que la calificación final concedida a un alumno/a en cualquier asignatura refleje de forma precisa su nivel de conocimiento y capacidad. Con el fin de lograr esto, se aplican los mismos estándares de evaluación para todos los alumnos/as, independientemente de si necesitan adecuaciones inclusivas de acceso.

Las adecuaciones inclusivas de acceso, incluidos los ajustes razonables, son medidas que se adoptan antes de un examen para que un alumno/a pueda acceder a la evaluación. Dichas adecuaciones no pueden solicitarse con carácter retroactivo, ya sea para exámenes orales, escritos o digitales.

Las adecuaciones de acceso solicitadas para un alumno/a no le deben representar una ventaja en ningún componente de evaluación.

Las adecuaciones inclusivas de acceso se han concebido para estudiantes que tengan un nivel de aptitud suficiente a fin de satisfacer todos los requisitos de evaluación.

Cuando un alumno/a necesita adecuaciones inclusivas de acceso, el colegio puede proporcionárselas basándose en los criterios del IB que se establecen claramente en la *Política de acceso e inclusión*. Las adecuaciones que se soliciten para la evaluación deben reflejar la forma habitual de trabajar del alumno/a. El colegio tiene la responsabilidad de seleccionar las adecuaciones que mejor se adapten a las necesidades del alumno/a a lo largo del programa de estudios, para garantizar una transición fluida a las evaluaciones sumativas. Las solicitudes de adecuaciones que no se ajusten a la forma de trabajo habitual del alumno/a solo se aprobarán en casos muy excepcionales e inusuales.

La filosofía educativa del IB se basa en la mentalidad internacional. La política es resultado de un estudio de las prácticas establecidas en diferentes países y regiones o territorios con el objeto de lograr la equidad entre estudiantes que tengan necesidades específicas de acceso a la evaluación.

Cada solicitud de adecuaciones inclusivas de acceso se juzgará por separado. Previas autorizaciones de adecuaciones, ya sean por parte del IB o de otro organismo de evaluación, no influirán en la decisión sobre la autorización de las adecuaciones que se hayan solicitado.

El IB trata toda la información sobre el alumnado de forma confidencial. Solo podrán acceder a dicha información, de ser necesario, el personal habilitado del IB y los/as miembros del Comité de la Evaluación Final del IB, a quienes se les pide que respeten su carácter confidencial.

Si un colegio no satisface las condiciones especificadas por el IB a la hora de administrar las adecuaciones inclusivas de acceso, o introduce adecuaciones especiales sin autorización, es posible que al alumno/a no se le conceda una calificación final en la asignatura y nivel en cuestión.

Si se puede demostrar que la falta de dominio de la lengua de respuesta de un alumno/a es consecuencia de una necesidad de apoyo para el aprendizaje que se haya detectado, se pueden autorizar adecuaciones inclusivas de acceso.

No se debe informar al personal examinador acerca de las circunstancias de un alumno/a ni de las condiciones adversas.

En el caso de los trabajos evaluados internamente, el profesorado no debe hacer ningún ajuste al corregir el trabajo del alumno/a.

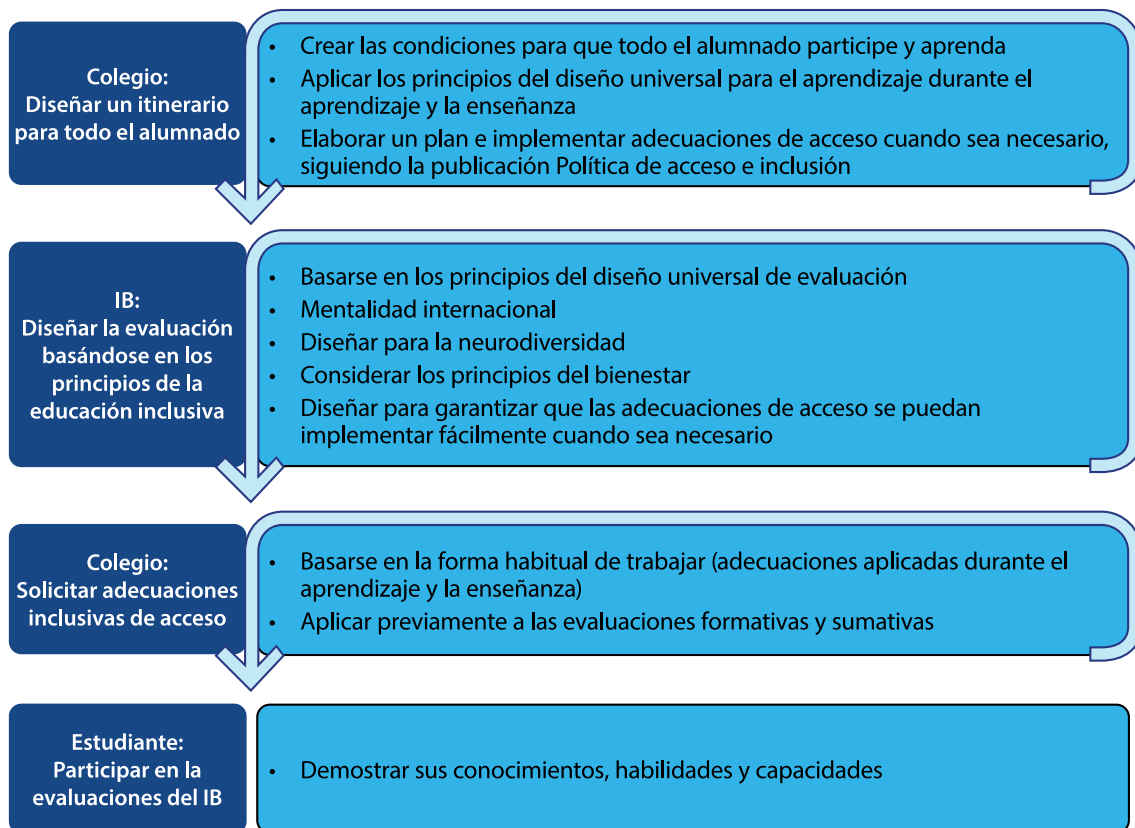
La lista de las adecuaciones inclusivas de acceso permitidas se revisa regularmente. Se pueden tomar en consideración adecuaciones alternativas que proponga un colegio, siempre y cuando todo el alumnado con necesidades similares pueda beneficiarse de dichas adecuaciones.

Si el alumno/a o la adecuaciones inclusivas de acceso que se hayan autorizado pudieran causar molestias a otros alumnos/as durante un examen, el alumno/a deberá realizar el examen en una sala independiente y bajo supervisión de acuerdo con el reglamento que rige la realización de los exámenes de cada programa.

Cualquier problema que surja como consecuencia de la naturaleza de las adecuaciones inclusivas de acceso o cualquier dificultad imprevista con la que se encuentre el alumno/a durante los exámenes deberán comunicarse a “El IB Responde” lo antes posible.

Figura 24

Implementación de adecuaciones inclusivas de acceso



Exenciones de evaluación

Normalmente no se autorizan exenciones respecto de los componentes de evaluación. No obstante, si un componente de evaluación, o una parte de un componente, exige una función fisiológica que un alumno/a no es capaz de realizar, se puede autorizar una exención. Antes de solicitar la exención respecto de un componente de evaluación, se deberán considerar detenidamente todos los ajustes razonables. Solo se autorizará una exención cuando existan motivos importantes. La incapacidad del alumno/a para realizar las funciones requeridas por el componente debe documentarse de forma clara y completa.

Para obtener información completa sobre los principios y procesos de las exenciones de evaluación, consulte la lista de adecuaciones inclusivas de acceso que se encuentra en la *Política de acceso e inclusión*.

Oportunidades para adecuaciones inclusivas de acceso con la evaluación digital

La evaluación digital permite que el alumnado tenga más control sobre el modo en que le gustaría que se presentara la evaluación. Los dispositivos informáticos ofrecen una amplia gama de fuentes, tamaños de texto y colores para satisfacer las preferencias y necesidades individuales, lo que hace que todo el alumnado pueda acceder a los ajustes. Aunque aplicar los principios de diseño universal mejora la accesibilidad, será necesario realizar algunas adecuaciones para facilitar el acceso. Para lograr esto, se tienen en cuenta funciones digitales que ofrecen opciones para implementar dichas adecuaciones. Las adecuaciones disponibles se enumeran en la [Política de acceso e inclusión](#).

El objetivo de las adecuaciones inclusivas de acceso en los exámenes digitales es eliminar barreras, al igual que en los exámenes en papel. Para garantizar la equidad, las adecuaciones de acceso aprobadas deben ser coherentes en los exámenes digitales y en papel.

Circunstancias adversas

Circunstancias adversas o imprevistas son todas aquellas circunstancias ajenas a la voluntad del alumno/a que puedan perjudicar su desempeño. Esto incluye enfermedad o lesión temporal, situaciones de estrés grave, circunstancias familiares excepcionalmente difíciles, pérdida de un ser querido o sucesos que puedan poner en peligro la salud o la seguridad del alumno/a. También se considera circunstancia adversa a cualquier acontecimiento que afecte a toda la comunidad escolar, como disturbios civiles o desastres naturales.

No se considerarán circunstancias adversas las deficiencias del colegio. Es responsabilidad del colegio que todo el alumnado cumpla los requisitos del programa y de la evaluación, incluidas las cuestiones que tienen que ver con el personal docente.

Para obtener información completa sobre qué se incluye y qué no en la categoría de circunstancias adversas y las medidas de mitigación que se ofrecen, consulte la [Política de circunstancias adversas](#).

Diseño universal de evaluación

En las secciones anteriores se ha analizado cómo se gestionan las situaciones que requieren adecuaciones inclusivas de acceso para garantizar un acceso equitativo a las evaluaciones del IB. Sin embargo, la solución ideal es diseñar desde el principio evaluaciones sin barreras para la participación. El concepto de diseño universal de evaluación hace hincapié en tener en cuenta el acceso, la inclusión, la equidad, la diversidad cultural y la sensibilidad. Siguiendo este enfoque, las evaluaciones del IB se elaboran de tal modo que se eliminen muchas barreras del modelo de evaluación general durante el proceso de diseño, desarrollo e implementación.

Crear desde el principio evaluaciones más inclusivas y con más pertinencia del constructo minimiza considerablemente la necesidad de iniciar o implementar adecuaciones, lo que beneficia a un mayor número de estudiantes.

El diseño universal de evaluación es un aspecto del marco más amplio de diseño universal para el aprendizaje, cuyo objetivo es crear entornos de aprendizaje accesibles para todo el alumnado. Este enfoque tiene en cuenta varios tipos de diversidad, como, por ejemplo, la neurodiversidad, las preferencias, las capacidades, los desafíos, los intereses, las culturas, las lenguas y los contextos.

Los principios del diseño universal para el aprendizaje se pueden aplicar a diversos aspectos de la educación, como el diseño del currículo y la evaluación, el desarrollo del currículo y la evaluación, y la gestión escolar. Para obtener más información sobre el diseño universal para el aprendizaje en el contexto del IB, consulte Rao *et al.* (2016).

El ciclo de vida de la evaluación

- El proceso de creación de las evaluaciones debe imaginarse como un ciclo continuo en el que cada etapa recibe la influencia de la anterior e influye en la siguiente.
- En promedio, se tarda 18 meses en crear un examen y su correspondiente esquema de calificación. Por este motivo, el IB los elabora en paralelo para las distintas convocatorias.

La evaluación es un ciclo con varias etapas: creación, organización de las convocatorias de exámenes para el alumnado, corrección por parte del personal examinador, y publicación de los resultados. Lo importante de todo este proceso es que, de cada convocatoria de exámenes, el IB obtiene lecciones para mejorar la calidad de la evaluación en la siguiente convocatoria.

La figura 25 es solo una posible ilustración de este proceso. Aunque muchas de las etapas se podrían separar o combinar de maneras diferentes, es una buena manera de explicar el ciclo de vida de una evaluación. En la tabla 6 se resume cada etapa y se ofrece un enlace a la sección pertinente.

Figura 25

El ciclo de evaluación

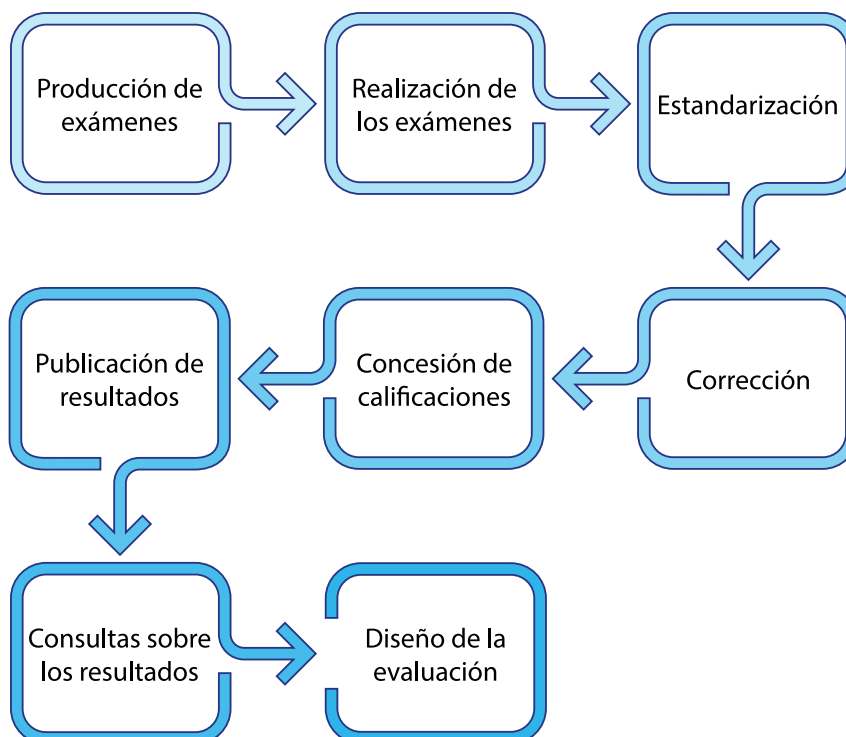


Tabla 6

Etapas del ciclo de vida de una evaluación

Etapa del ciclo de evaluación	Descripción
Concesión de calificaciones	Etapa en la que los examinadores/as supervisores deciden cómo convertir las puntuaciones (que dependen del examen) en calificaciones finales (que siempre significan lo mismo, independientemente del examen). Las actividades posteriores a la concesión de calificaciones incluyen la supervisión de la distribución de

Etapa del ciclo de evaluación	Descripción
	calificaciones de cada examinador/a y la identificación de estudiantes que corren el riesgo de obtener una calificación incorrecta, basándose en pruebas de que ha habido un problema concreto.
Consultas sobre los resultados	Posibilidad que se ofrece a los colegios para indicar al IB los errores que consideren que ha habido en el proceso de corrección de exámenes y para pedir que se revise el trabajo de un alumno/a.
Corrección	Parte en la que el personal examinador revisa el trabajo de cada estudiante y decide qué puntuación le otorga. Los examinadores/as deben seguir las instrucciones del examinador/a principal, y la calidad de sus correcciones se comprueba mediante controles regulares.
Diseño de la evaluación	El desempeño del alumnado y su interpretación de las preguntas de examen sirven de base para el diseño de las evaluaciones. El IB utiliza esta información para mejorar lo que se pretende evaluar y el enfoque que adopta, incluidos el número y el tipo de tareas.
Estandarización	Proceso en el que los examinadores/as supervisores explican a su equipo examinador cómo corregir los trabajos del alumnado, e identifican los exámenes que están corregidos definitivamente y que se utilizarán en los procesos de garantía de calidad de la corrección.
Producción de exámenes	Proceso de creación de cada examen. Cubre un amplio espectro: decidir los temas que tratarán las preguntas, redactar y corregir las preguntas concretas, traducir el examen a otras lenguas, darle el formato deseado y, finalmente, realizar los controles de calidad necesarios.
Publicación de resultados	Difusión de los resultados para los colegios y el alumnado.
Realización de los exámenes	Parte del proceso en la que el alumnado completa las evaluaciones en los colegios.

Para obtener más información sobre el ciclo de vida de la evaluación, consulte el video [“El ciclo de vida de un examen: Los pasos seguidos desde que se crean los exámenes hasta que se otorgan las calificaciones”](#).

El impacto de la evaluación digital en el ciclo de evaluación

La introducción de la evaluación digital no supone ningún cambio para el principio del ciclo de evaluación. Ahora bien, la evaluación electrónica permitirá ir más rápido en algunas partes del ciclo (por ejemplo, ya no hará falta enviar los cuestionarios de examen en papel y los exámenes escritos del alumnado a centros de escaneado). A pesar de todo, habrá que seguir completando cada una de las partes del ciclo de evaluación.

Elaboración de evaluaciones

- Las pruebas de evaluación deben pasar por distintas etapas de finalización durante la producción antes de llegar al alumnado:
 - Creación de evaluaciones y aprobación del contenido
 - Formato y revisión
 - Usabilidad
 - Traducción
 - Envío a los colegios
- En esta etapa también se tienen en cuenta las modificaciones que solicite el alumnado (adaptaciones inclusivas de acceso) durante el proceso de finalización.
- Independientemente de que los exámenes sean en papel o digitales, los principios del ciclo de elaboración de evaluaciones no cambian.

Componentes de una buena evaluación

Además de las cualidades que caracterizan a una buena evaluación descritas en la “Sección A: Principios de evaluación”, el IB espera que los/as responsables de desarrollo del currículo y los equipos de creación de evaluaciones tengan en cuenta los siguientes elementos clave:

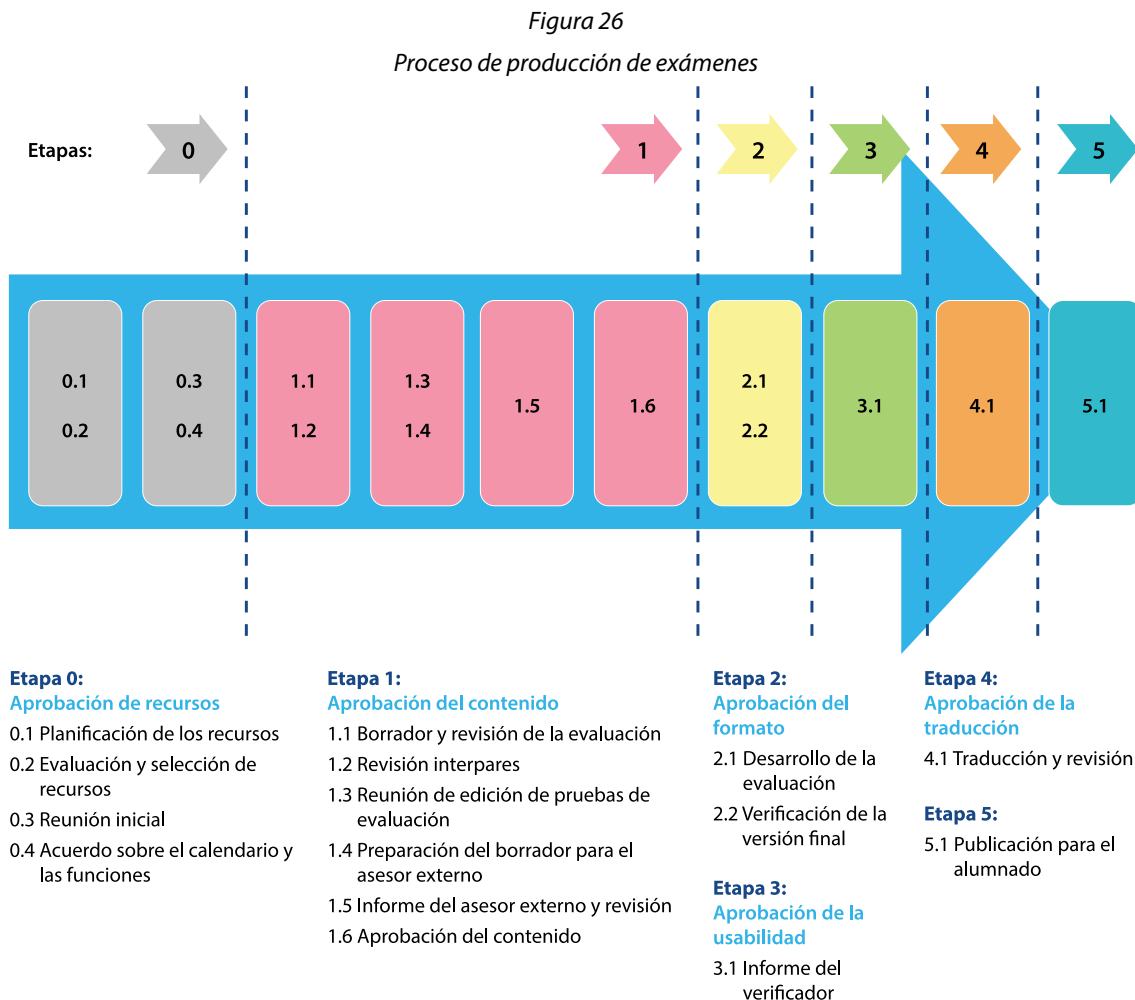
- La evaluación debe proporcionar al alumnado oportunidades para que demuestre lo que es capaz de hacer, en vez de servir para identificar lo que no sabe hacer.
- Las evaluaciones deben ser accesibles para la mayor cantidad posible de estudiantes, pero también deben poder permitir una diferenciación razonable.
- Las mejores evaluaciones son las que resultan accesibles para todo el alumnado. El IB utiliza el diseño universal en toda la elaboración de evaluaciones para reducir, en la medida de lo posible, la necesidad de adoptar adaptaciones inclusivas de acceso.
- La evaluación debe tratar solamente sobre el currículo tal como aparece en la guía de la asignatura.
- Al desarrollar la evaluación hay que plantearse cómo se corregirá, con el fin de que se concedan puntos a los aspectos que se quieren evaluar. Esto significa que el esquema de calificación se desarrolla al mismo tiempo que los exámenes, y que se incluye orientación para los examinadores/as sobre los márgenes de tolerancia previstos.

Términos de instrucción

Los términos de instrucción son palabras y expresiones clave utilizadas en los contenidos del programa de estudios y en las preguntas de examen para indicar lo que se pide en respuesta a una instrucción dada, así como el tipo y la profundidad de la respuesta que se espera del alumnado.

Los términos de instrucción plantean un desafío al traducirlos a otras lenguas, porque las sutilezas del significado con frecuencia pueden perderse por el camino. Para evitar que esto suceda, el IB publica sus términos de instrucción en las guías de las asignaturas en todas las lenguas de respuesta, y explica su significado. Al revisar los exámenes traducidos, una de las tareas clave es comprobar que se utilicen los términos de instrucción correctos.

Descripción general de la producción de exámenes



La figura 26 ilustra las distintas etapas del proceso de producción de exámenes. La creación de un examen es un proceso largo de unos 18 meses de duración e implica la participación de una amplia variedad de personal experto externo y personal del IB, como autores/as, responsables de asignaturas, especialistas en la asignatura, editores/as, verificadores/as externos y lingüistas.

Es muy importante desarrollar el esquema de calificación al mismo tiempo que el examen, para garantizar que las preguntas y las puntuaciones vayan en consonancia con lo que pretenden evaluar.

El proceso de aprobación del contenido

La primera etapa del proceso es determinar a las personas que se encargarán de redactar los exámenes. Normalmente esto se encarga a los examinadores/as principales, pero después de comprobar cualquier factor que indique que podría haber un conflicto de intereses, como las relaciones con los colegios o la responsabilidad de realizar talleres de capacitación.

En la producción de exámenes intervienen:

- Autor/a (miembro externo)
- Asesor/a externo o normalizador/a (miembro externo)
- Verificador/a o persona con buen dominio de la lengua (miembro externo)
- Responsable de la asignatura (miembro del personal del IB)
- Editor/a de producción (miembro del personal del IB)
- Corrector/a (miembro del personal del IB)

- Diseñador/a (miembro del personal del IB)
- Traductor/a (miembro externo)
- Revisor/a de la traducción (miembro externo)

Cuando el IB prepara exámenes para distintas zonas horarias, los exámenes se someten a procesos de producción totalmente independientes.

Creación y aprobación del contenido

Los autores/as deben elaborar un primer borrador del examen y del esquema de calificación. Estos borradores se someten a varios niveles de revisión con otros autores/as y responsables de asignaturas. En algunas asignaturas, más tarde se celebra una reunión de edición de pruebas de evaluación, en la que cada examen se revisa para comprobar lo siguiente:

- Nivel de exigencia cognitiva de las preguntas y el grado de dificultad general del examen
- Posibles sesgos
- Accesibilidad
- Cobertura del currículo
- Correspondencia entre el examen y el modelo de evaluación publicado
- Relación del nuevo examen con exámenes y materiales de muestra anteriores
- Verificación de que las preguntas no sean iguales a las que se usaron como ejemplo en algún recurso publicado

Esta reunión da como resultado otro conjunto de borradores que posteriormente revisará el asesor/a externo o el normalizador/a.

El asesor/a externo o el normalizador/a es una persona experta en la asignatura que aún no ha participado en la redacción de los exámenes ni de los esquemas de calificación. Esta persona aporta una nueva perspectiva a lo que se ha creado hasta el momento, y debe comentar el nivel de dificultad de los exámenes en conjunto, cómo son en comparación con exámenes anteriores, y en qué medida se corresponden con el currículo. El asesor/a externo propone cambios que deberán revisar el autor/a y la persona responsable de la asignatura para luego escribir las versiones finales.

A continuación, se realiza una revisión interna de los borradores finales, con comprobaciones por parte de un corrector/a y un diseñador/a, en colaboración con la persona responsable de la asignatura y el autor/a, si es necesario.

El autor/a es la persona encargada de dar la aprobación al contenido final de la prueba de examen, y la persona responsable de la asignatura del IB es quien confirma que el examen se ha revisado adecuadamente y respetando los estándares de calidad establecidos del IB.

El esquema de calificación no se aprueba en esta fase, sino que se revisa durante la estandarización en función de las respuestas del alumnado.

Formato y revisión

Una vez aprobado el contenido, el texto de las preguntas no se puede modificar. A continuación, el diseñador/a dará a las preguntas el formato que verá el alumnado en el examen (ya sea en papel o digital).

En los exámenes tradicionales (en papel), esto implica añadir la portada con las instrucciones, aplicar las normas de estilo adecuadas, y añadir los códigos de barras y las páginas en blanco para dar forma al cuadernillo de preguntas. Este proceso también conlleva volver a realizar algunos diagramas para que tengan la calidad necesaria, o buscar imágenes adecuadas en línea.

Para la evaluación digital, el diseñador/a tendrá que crear los exámenes en un entorno específico. En él se incluyen las instrucciones y las guías de estilo adecuadas.

En ambos casos, las pruebas de examen completadas se comparan con el contenido aprobado. Esta labor la realiza un editor/a de producción.

Aprobación de la usabilidad

El verificador/a es un experto en la asignatura que aún no ha participado en el proceso de producción de los exámenes. Esta persona realiza la evaluación como si fuera un alumno/a, y toma nota del tiempo que ha necesitado para responder las preguntas. El propósito de este control es detectar posibles errores e identificar si hay alguna ambigüedad en las instrucciones o en las preguntas.

Además, el verificador/a revisa el esquema de calificación una vez que ha realizado la evaluación.

En algunas asignaturas de lengua con poco alumnado, la función del verificador/a la realiza un segundo hablante nativo de la lengua en vez de una persona experta en literatura.

Después de considerar los comentarios del verificador/a, las pruebas de evaluación se finalizan y están listas para ponerse a disposición de los colegios.

Control de calidad

Es muy importante que en las pruebas de examen no haya errores ni ambigüedades, ya que dichos problemas podrían tener repercusiones considerables para el alumnado que realiza los exámenes. El IB se toma este tema muy seriamente.

En cada etapa del ciclo de elaboración de evaluaciones existe un control de la calidad, además de la aprobación oficial.

Traducciones

El IB da al alumnado la oportunidad de que los exámenes (que no sean de Lengua) se realicen en distintas lenguas de respuesta. Actualmente, la mayoría de las asignaturas están disponibles en español, francés e inglés. En ocasiones, y debido a acuerdos locales, algunas asignaturas también se ofrecen en alemán, coreano y japonés.

La necesidad de traducir los exámenes se tiene en cuenta desde los primeros borradores, y los equipos del IB encargados de la creación y revisión de pruebas de examen son muy conscientes de los problemas que pueden surgir. El proceso formal de traducción comienza una vez finalizada la elaboración de los exámenes en inglés.

Es fundamental que el proceso de traducción no altere el significado de las preguntas, para evitar que haya alumnos/as que salgan favorecidos o desfavorecidos al realizar la evaluación en una lengua concreta. Los traductores/as externos que contrata el IB son especialistas en la asignatura, para garantizar que el lenguaje técnico que se utiliza en la evaluación sea preciso. Una vez entregada la traducción, un revisor/a bilingüe la compara con el examen original (y no solamente el texto) para hacer un último control de calidad.

Pruebas de examen modificadas

Lo ideal es diseñar las evaluaciones de modo que resulten adecuadas para todo el alumnado. Sin embargo, hay necesidades y requisitos que no se pueden satisfacer mediante la incorporación del diseño universal, por lo que será necesario proporcionar versiones modificadas. Estas se gestionan de acuerdo con la [Política de acceso e inclusión](#), y se elaboran y se entregan a los colegios de la misma manera que las evaluaciones no modificadas.

Exámenes

- El propósito de los exámenes es permitir que el alumnado demuestre conocimientos, comprensión y aplicación de habilidades en un entorno controlado que ofrezca una experiencia coherente a todos los alumnos/as y que minimice los riesgos de conducta impropia.
- Durante la preparación de las convocatorias de exámenes, el IB tiene que hallar un equilibrio entre las necesidades de todo el alumnado y la manejabilidad para los colegios, y finalizar la corrección con la máxima rapidez para cumplir con las expectativas de los alumnos/as.

- El IB publica reglas claras sobre el comportamiento en el lugar de realización del examen para minimizar las ocasiones de que se produzcan conductas improcedentes. Estas reglas se actualizan en función de los nuevos avances tecnológicos y los cambios en los entornos de trabajo.
- El calendario de exámenes intenta responder de manera equilibrada a las numerosas prioridades opuestas que existen, y representa “la opción menos mala” a nivel internacional.
- En los *Procedimientos de evaluación* del programa correspondiente se detalla toda la información relacionada con los hechos inesperados, las adecuaciones inclusivas de acceso, las circunstancias adversas y los cambios de fechas.

La época de exámenes puede ser muy estresante y exigente para los alumnos/as. Con el fin de reducir la presión a la que se ve sometido el alumnado, se deben respetar los siguientes principios:

- Limitar la duración de la convocatoria de exámenes
- Evitar, en la medida de lo posible, el solapamiento con otros exámenes (por ejemplo, exámenes nacionales)

Es importante encontrar un equilibrio entre estas medidas, la necesidad de disponer de suficientes datos de evaluación para extraer conclusiones válidas sobre el desempeño del alumno/a, y evitar que haya conductas improcedentes.

Preparación y gestión del lugar de realización del examen

Los colegios deben realizar los exámenes siguiendo unas reglas muy estrictas estipuladas en la publicación *Instrucciones sobre la realización de los exámenes* del programa correspondiente.

Mantener la seguridad de los exámenes y una conducta adecuada en los exámenes es de gran importancia, ya que si los colegios u otras partes interesadas tienen la impresión de que ha habido algún caso de conducta improcedente, el valor de los resultados del alumnado disminuirá considerablemente. El IB tiene establecida una serie de políticas y procedimientos para mantener la confianza general, tal como se explica en la sección “Integridad de la evaluación” de esta publicación.

Inevitablemente, el IB tiene que abordar incidentes en los que no dispone del trabajo de un alumno/a y no lo puede corregir, por motivos que escapan del control del colegio y del alumno/a. Para intentar mitigar los perjuicios, el IB puede hacer una estimación de la puntuación que se le concede al alumno/a, aunque esto solo es posible si hay otras pruebas en las que basarse. Para asegurarse de que así sea en el caso de los exámenes en papel, el IB requiere que los distintos cuestionarios de examen con las respuestas del alumnado se envíen al centro de escaneado en días diferentes, ya que así se reduce la posibilidad de que todos se pierdan por el camino.

La evaluación digital disminuye este riesgo en el contexto del PD y del POP, ya que se dispondrá de copias del trabajo entregado para evaluación en cada parte del proceso de carga.

Desarrollo del calendario de exámenes

Con el fin de que las convocatorias de exámenes sean lo más breve posible, el IB hace coincidir distintas asignaturas a la misma hora en el calendario de exámenes. Aunque se tienen en cuenta las posibles combinaciones de asignaturas del alumnado, en ocasiones es inevitable que haya una pequeña cantidad de solapamientos. El procedimiento que debe seguir en esos casos se explica en los *Procedimientos de evaluación* del programa correspondiente.

Los exámenes se programan para evitar, siempre que sea posible, que haya más de seis horas y media de exámenes en un día. En general, las dos o tres pruebas de un examen con varios componentes se realizan consecutivamente, de forma que la primera prueba se realiza una tarde y la última se realiza la mañana del día siguiente. Aunque esto no siempre es posible, esta distribución es preferible a la de organizar todos los exámenes de un curso determinado el mismo día.

Para la mayoría del alumnado, esto supone una distribución más uniforme de la carga de trabajo durante las dos semanas (en el caso del PAI) o las tres semanas (en el caso del PD y del POP) del calendario de exámenes. Esto permite al alumnado recuperarse durante la noche en caso de que crea que no ha dado todo lo que puede de sí mismo en un examen concreto.

Principios del diseño del calendario de exámenes

Los puntos que se indican a continuación son un resumen de los principios en los que se basa la creación del calendario de exámenes.

No siempre es posible cumplir todos estos principios, así que a veces es necesario llegar a un punto medio. El IB publica el calendario de exámenes al menos un año antes de que se celebren.

- Dado el elevado número de países y regiones o territorios en los que se ofrecen los programas del IB, no es posible tener en cuenta los días festivos públicos, nacionales o de cada colegio (excepto el 1 de mayo).
- A pesar de que lo ideal sería que los exámenes no se celebraran los jueves ni los viernes, por respeto a los colegios de Oriente Medio —cuyo fin de semana es en esos días—, actualmente en la práctica esto no es posible.
- El IB tendrá en cuenta los casos en los que haya un estrecho vínculo entre una asignatura y una región o cultura en particular. Por ejemplo, los exámenes de literatura o lengua árabe no se programarán para un viernes.
- El IB utiliza los datos de matriculación relativos a las combinaciones de asignaturas para garantizar que la coincidencia de asignaturas en el calendario afecte a la menor cantidad de estudiantes posible en todo el mundo.
- No se examinará al alumnado en dos lenguas extranjeras distintas en el mismo día (esta medida no concierne a las lenguas de respuesta).
- A fin de minimizar el riesgo de que suceda algún acontecimiento inesperado que ponga en desventaja a un alumno/a en todos los componentes de los cursos en los que hay varios exámenes, dichos exámenes se programarán para realizarse en dos días como mínimo.
- El alumnado debe poder concentrar su repaso de cada asignatura en un plazo de tiempo limitado. Por lo tanto, siempre que sea posible, aquellos exámenes que se celebren en más de un día se intentarán programar para días consecutivos.
- Los exámenes se deben repartir durante todo el período de exámenes, y no concentrarse en un período más corto de tiempo. Esto significa que, cuando sea posible, los exámenes de las asignaturas con un mayor número de estudiantes (como Historia, Inglés o Matemáticas) no se programarán para días consecutivos.
- Por ese mismo motivo, el IB intentará programar los exámenes de Lengua y de Ciencias para cada una de las dos semanas (en el caso del PAI) o las tres semanas (en el caso del PD y del POP) distintas del calendario de exámenes.
- Con el fin de poder satisfacer los requisitos internos para procesar los exámenes del IB, los exámenes de algunas asignaturas normalmente se celebrarán al principio del calendario. En particular, este será el caso de las asignaturas con un gran número de estudiantes.

Gestión de acontecimientos inesperados

En los *Procedimientos de evaluación* del programa correspondiente se explica cómo gestionar la mayoría de los acontecimientos inesperados. Si tiene dudas, comuníquese con [“El IB Responde”](#) para recibir orientación.

Adecuaciones inclusivas de acceso y circunstancias adversas

Estos temas se tratan con más detalle en la sección “[Priorización de la equidad para todo el alumnado](#)” de esta publicación.

El IB cree que todo el alumnado debe tener la posibilidad de demostrar sus capacidades en las condiciones de evaluación más equitativas que sean posibles. Las condiciones normales de evaluación pueden representar una desventaja para estudiantes con necesidades de apoyo para el aprendizaje, al impedirles demostrar su nivel de conocimiento y capacidad. En estos casos, se pueden autorizar adecuaciones inclusivas de acceso, que pueden ser disposiciones especiales para realizar el examen o modificaciones de los exámenes.

Se denomina *circunstancia adversa* a cualquier acontecimiento ajeno a la voluntad del alumno/a que ocurre durante la convocatoria de exámenes y que puede perjudicar su desempeño. Esto incluye enfermedad o lesión temporal, situaciones de estrés grave, circunstancias familiares excepcionalmente difíciles, pérdida de un ser querido o sucesos que puedan poner en peligro la salud o la seguridad del alumno/a. También se considera circunstancia adversa a cualquier acontecimiento que afecte a toda la comunidad escolar, como disturbios civiles o desastres naturales. No se considerarán circunstancias adversas las deficiencias del colegio.

Para obtener más información, consulte los *Procedimientos de evaluación* del programa correspondiente.

Cambio de fechas de los exámenes

Cambiar la fecha de un examen plantea riesgos importantes para la integridad de dicho examen. El motivo es que algunos alumnos/as realizarían el examen antes o después que otros/as, lo cual fácilmente podría conducir a alguna conducta impropia. En ningún caso se permite cambiar un examen al día anterior en que estaba previsto, porque el riesgo de cometer una infracción que afectase a la mayoría del alumnado que realice el examen sería demasiado grande.

Las únicas circunstancias en las que el IB permitirá a un alumno/a realizar uno o varios exámenes en una hora o fecha distinta a la programada, y los procesos para solicitarlo, se definen detalladamente en los *Procedimientos de evaluación* del programa correspondiente y en la publicación *Política de preparación para los exámenes*, ambos disponibles en el [Centro de Recursos para los Programas](#).

Corrección

- Se considera que la corrección se ha realizado satisfactoriamente cuando se da al alumnado una puntuación coherente y precisa que refleja la calidad de su trabajo.
- Se considera que las indicaciones del examinador/a principal son correctas, y los demás examinadores/as deben seguirlas cuando corrijan.
- Puntuación y calificación son conceptos distintos: es posible que el alumnado obtenga una mayor puntuación en un examen más fácil, pero que se le conceda la misma calificación.

Puntuación y calificación

Un aspecto importante en el desarrollo y la utilización de las evaluaciones sumativas del alumnado es comprender la diferencia entre corregir y calificar su trabajo.

- Al corregir, se reconoce la labor que ha realizado un alumno/a y se le otorgan unas puntuaciones según un esquema de calificación o un marco similar. Esto indica en qué medida el alumno/a realizó correctamente la tarea de evaluación. La puntuación en sí misma no significa nada más.
- A la hora de decidir la calificación, el examinador/a juzga la calidad del trabajo del alumno/a utilizando como referencia un estándar definido. Para ello, se tienen en cuenta la dificultad de la tarea y también la proporción de ella que se completó. Así pues, la calificación tiene cierto significado o pertinencia, y en principio debe poder compararse con el desempeño logrado en otras evaluaciones.

Es posible recibir una calificación alta al responder bien solamente una parte pequeña de una pregunta muy difícil. Del mismo modo, puede resultar imposible conseguir la misma calificación si solo se responden correctamente muchas preguntas triviales.

Tal como se indica en secciones posteriores, no es necesario hacer referencia a los logros del alumno/a para explicar el nivel correspondiente a la calificación, si bien el IB adopta ese enfoque. Existen otros sistemas perfectamente coherentes y muy respetados en los que los niveles de logro se basan en el desempeño del alumno/a en comparación con sus compañeros/as.

En sus evaluaciones, el IB normalmente utiliza la puntuación como indicativo del desempeño general. Posteriormente, estudia el desempeño del alumnado que ha obtenido esa puntuación para determinar una puntuación límite (límite de calificación) en la que se concede una calificación concreta al alumnado que supera una puntuación determinada. Este proceso se explica con más detalle en la sección [“Concesión de calificaciones finales y totalización”](#).

Enfoques de la corrección

Dado que el IB utiliza diferentes instrumentos de evaluación, es esencial seleccionar el enfoque de corrección adecuado para cada evaluación. La decisión sobre el enfoque de corrección se toma durante el proceso de desarrollo del currículo.

Esquemas de calificación analíticos

Los esquemas de calificación analíticos se utilizan para aquellas evaluaciones en las que se espera que el alumnado conteste con una variedad limitada de respuestas.

Estos esquemas de calificación pueden dar instrucciones específicas a los examinadores/as sobre cómo repartir la puntuación total disponible para una pregunta entre las diferentes partes de la respuesta. Incluso en el caso de las preguntas estructuradas en las que se espera recibir respuestas muy específicas, los esquemas de calificación deben proporcionar a los examinadores/as suficiente información para poder

corregir de manera coherente, teniendo en cuenta los diferentes tipos de enfoques que podría adoptar el alumno/a, así como los errores frecuentes que podría cometer. Inevitablemente, los examinadores/as tendrán que utilizar su juicio profesional para asignar puntuaciones a las respuestas inesperadas o a otras respuestas válidas, pero los esquemas de calificación deberán aportarles la mejor orientación posible sobre cómo tomar esas decisiones.

No siempre es posible predecir qué partes de una pregunta responderá bien o mal el alumnado. Esto resulta especialmente delicado en las preguntas estructuradas de respuesta larga, ya que los errores en una parte inicial pueden tener repercusiones más adelante. El IB diseña este tipo de preguntas de tal modo que si un alumno/a comete un error en la parte inicial de una pregunta, no se anule la validez que pueda tener el resto de la respuesta. No obstante, los esquemas de calificación analíticos deben ofrecer una orientación explícita sobre cómo corregir determinados tipos de respuestas incorrectas y cómo seguir con el resto del trabajo del alumno/a cuando haya cometido un error en una parte de una pregunta.

Criterios de evaluación

En el caso de otras evaluaciones en las que un esquema de calificación detallado no sería adecuado, se aplican en su lugar criterios de evaluación.

Los criterios de evaluación se centran más en el tipo de desempeño que se espera que demuestre el alumnado, sin importar los detalles concretos que se aporten en cada respuesta. Con objeto de reflejar los distintos niveles de conocimientos que se pueden demostrar, existen los descriptores de nivel.

En el IB, los descriptores son elementos fundamentales para determinar qué nivel de logro alcanza cada respuesta de un alumno/a. Basándose en los descriptores, el examinador/a que aplica un criterio de evaluación debe elegir el nivel de logro que mejor se corresponda con el trabajo que está corrigiendo. No es necesario cumplir todos y cada uno de los aspectos de un nivel de logro para alcanzar ese nivel. Además, hay que recordar que el nivel más alto de cualquiera de los criterios no describe un trabajo perfecto.

En la mayoría de los componentes que se corrigen utilizando criterios de evaluación, se utilizan varios criterios, en lugar de uno solo. Es importante que los criterios sean independientes los unos de los otros. No sería justo que un alumno/a recibiera puntos en distintas partes por el mismo aspecto de su respuesta.

Es habitual reutilizar los mismos criterios de evaluación cada año, aunque cambien las preguntas que se plantean al alumnado. Esto se debe a que la naturaleza subyacente de lo que se evalúa no varía. Las preguntas nuevas normalmente tratan sobre detalles específicos que no figuran explícitamente en los criterios. El examinador/a principal suele proporcionar notas para la corrección que sirven para explicar al personal examinador cómo aplicar los criterios de evaluación en cada pregunta, y es posible que también proporcione ejemplos de detalles pertinentes.

Criterios holísticos: bandas de puntuación

En ocasiones, no es adecuado separar los distintos criterios de evaluación al corregir un trabajo. Esto sucede normalmente cuando es imposible tener distintos criterios que sean independientes entre sí. En esos casos, se utilizan bandas de puntuación en vez de criterios separados. En efecto, las bandas de puntuación representan un criterio holístico único que se aplica al trabajo, el cual se juzga en su totalidad. Dado que se necesita una variedad razonable de puntuaciones para poder reflejar los distintos niveles de desempeño del alumnado, cada descriptor de nivel de la banda de puntuación corresponde a un rango de puntuaciones diferentes.

Tabla 7

Ejemplo de bandas de puntuación del IB

Puntuación	Descriptor de nivel
0	El trabajo no alcanza ninguno de los niveles especificados por los descriptores que figuran a continuación.
1-3	<ul style="list-style-type: none"> Se demuestra una escasa comprensión de la cuestión. No se utiliza vocabulario específico de la asignatura, o se utiliza constantemente de manera inapropiada.

Puntuación	Descriptor de nivel
	<ul style="list-style-type: none"> La explicación de la cuestión es mínima. Los argumentos aportados son superficiales y con frecuencia poco claros. La respuesta es descriptiva. Cuando se incluye análisis, este es superficial o incoherente. Hay poca discusión, o ninguna en absoluto, de distintos puntos de vista. Cuando se incluye una conclusión, esta es muy superficial o no es coherente con el resto de la respuesta.
4-6	<ul style="list-style-type: none"> Se demuestra una comprensión básica de la cuestión. Se utiliza vocabulario específico de la asignatura, pero a menudo de manera inapropiada. La explicación de la cuestión es básica y está poco desarrollada. Los argumentos son a menudo imprecisos o vagos, y con frecuencia no queda claro lo que se trata de decir en la respuesta. Hay un análisis limitado y la respuesta es, en general, más descriptiva que analítica. Hay una discusión limitada de distintos puntos de vista. La conclusión incluida es simplista.
7-9	<ul style="list-style-type: none"> Se demuestra cierta comprensión de la cuestión. Se utiliza vocabulario específico de la asignatura y, a veces, de manera apropiada. Se ofrece una explicación satisfactoria de la cuestión, pero le falta claridad y desarrollo en algunas partes. Se aportan argumentos pertinentes, pero les falta precisión y detalle. La respuesta incluye análisis, pero este carece de desarrollo. Hay cierta discusión de distintos puntos de vista. Se incluye una conclusión.
10-12	<ul style="list-style-type: none"> Se demuestra una buena comprensión de la cuestión. Se utiliza vocabulario específico de la asignatura y, en general, de manera apropiada. La explicación de la cuestión es clara, aunque requiere un mayor desarrollo. Se aportan argumentos pertinentes y precisos, pero les falta detalle. La respuesta incluye un análisis crítico, pero este carece de desarrollo. Hay discusión de distintos puntos de vista. La respuesta brinda argumentos que conducen a una conclusión, y esta es coherente con los argumentos presentados.
13-15	<ul style="list-style-type: none"> Se demuestra una muy buena comprensión de la cuestión. Se utiliza vocabulario específico de la asignatura de forma correcta y precisa. La explicación de la cuestión es clara y está bien desarrollada. Se aportan argumentos pertinentes, precisos y detallados. La respuesta incluye un análisis crítico bien desarrollado. Hay discusión crítica de distintos puntos de vista. La respuesta brinda argumentos que conducen a una conclusión razonada y formulada con claridad, y esta es coherente con los argumentos presentados.

En general, los descriptores suelen ser bastante largos porque cubren una variedad de cualidades que pueden aparecer en el trabajo del alumnado, y se relacionan directamente con los objetivos del curso. Tal como sucede con los criterios de evaluación, los descriptores se emplean para determinar qué nivel de logro alcanza cada respuesta de un alumno/a. Los examinadores/as deben decidir qué puntuación concreta conceder del rango de puntos disponibles en cada descriptor, para lo cual deben determinar en qué medida el trabajo del alumno/a se ajusta a dicho descriptor. Por ejemplo, un nivel de una banda de puntuación puede cubrir el rango de 7 a 9 puntos. El examinador/a concederá al trabajo del alumno/a la puntuación de ese rango que mejor se ajuste al descriptor de nivel pertinente.

Corrección automática

Una pequeña cantidad de componentes que consisten en preguntas de opción múltiple se corrigen automáticamente. La corrección automática es el proceso de utilizar la tecnología para evaluar el trabajo del alumnado según un esquema de calificación predefinido. En estas preguntas de opción múltiple, el esquema de calificación indica si una respuesta es objetivamente correcta o incorrecta. Los esquemas de calificación para estos componentes son muy claros y no es necesaria la valoración del personal examinador.

Correcciones que requieren la valoración del personal examinador

Aunque algunas evaluaciones pueden corregirse automáticamente, la mayoría de las evaluaciones del IB las debe corregir el personal examinador, ya que es necesario aplicar esquemas de calificación analíticos o criterios de corrección.

El personal examinador corrige de manera electrónica el trabajo del alumnado mediante un tipo de software denominado *herramienta de corrección*. Este software permite añadir signos de visto, puntuaciones y anotaciones en la copia electrónica del trabajo del alumno/a, incluso en materiales audiovisuales. Esto amplía considerablemente la variedad de trabajos que se pueden entregar para demostrar conocimientos y habilidades como, por ejemplo, grabaciones de representaciones o de exámenes orales.

Para obtener más información sobre cómo corrige el IB, consulte el video [“Los dos métodos para la corrección: cómo realizan la corrección los examinadores y los profesores”](#).

Estandarización

El proceso de estandarización cubre varios elementos del procedimiento de evaluación. El propósito de la estandarización es preparar al personal examinador para corregir con fiabilidad.

Para que los resultados sean fiables, es importante que cada examen reciba la misma puntuación en función de la calidad del trabajo que contiene, independientemente del examinador/a que lo corrija. Para ello, el personal examinador debe demostrar que comprende el estándar de corrección que establezca el examinador/a principal para un componente, y ajustarse a él, antes de poder corregir en una convocatoria real de exámenes del IB.

El primer paso del proceso de estandarización consiste en seleccionar a examinadores/as con las habilidades y la experiencia necesarias para corregir. A continuación, se les ofrece una capacitación para que entiendan los principios y las prácticas de la corrección.

Una vez que el alumnado ha realizado un examen, el equipo de examinadores/as supervisores, dirigido por el examinador/a principal, se reúne para discutir el esquema de calificación y revisar los exámenes que servirán de muestra. El objetivo de esta reunión de estandarización es garantizar que el esquema de calificación cubra adecuadamente la variedad prevista de respuestas que los examinadores/as pueden encontrar. Es poco probable que un esquema de calificación cubra todas las respuestas posibles que un alumno/a pueda dar, pero revisar en la reunión los exámenes de muestra permite al examinador/a principal y al equipo de examinadores/as supervisores elaborar instrucciones y orientación claras para informar al personal examinador sobre el tipo de respuesta requerida para cada puntuación, basándose en el trabajo visto hasta el momento.

El resultado de la reunión de estandarización es un conjunto de exámenes con las respuestas del alumnado que ha corregido de forma definitiva el examinador/a principal. Dichos exámenes se utilizan para capacitar a otros examinadores/as para corregir empleando el mismo estándar que el examinador/a principal. Solo a los examinadores/as que hayan demostrado seguir el estándar de corrección del examinador/a principal se les permite corregir los exámenes en una convocatoria real, y su estándar de corrección se comprueba repetidamente a lo largo del período de corrección con sólidos procedimientos de garantía de calidad.

El modelo de calidad

Es importante que el alumno/a reciba el mismo resultado que habría obtenido si un examinador/a principal hubiera realizado la corrección, independientemente de quién sea el examinador/a que corrija su trabajo.

Después de realizar la evaluación y de recibir algunas respuestas del alumnado, el examinador/a principal realiza una reunión de estandarización con los demás examinadores/as supervisores. Los examinadores/as supervisores deben acordar las puntuaciones para varias respuestas que se usarán para definir el modelo de calidad. Hay tres categorías de exámenes corregidos definitivamente: respuestas de práctica, respuestas de aptitud, y respuestas de control. Juntas forman el modelo de calidad.

Respuestas de práctica

El propósito de las respuestas de práctica es ayudar al personal examinador a aprender el estándar de corrección que ha establecido el examinador/a principal. Dichas respuestas deben mostrar cómo corregir ejemplos habituales de exámenes con respuestas del alumnado, e identificar cualquier situación común en la que pueda ser necesario un conocimiento específico del esquema de calificación.

Un buen conjunto de respuestas de práctica debe incluir un rango de ejemplos de desempeño del alumnado, con puntuaciones de vayan de bajas a altas, para ayudar a los examinadores/as a determinar qué puntuación asignar en cada caso y qué se entiende por una buena (o mala) respuesta.

Después de revisar el esquema de calificación o los criterios de corrección, y cualquier nota para la corrección que el examinador/a principal haya proporcionado después de la reunión de estandarización, los examinadores/as pueden pasar a la etapa de práctica. En las respuestas de práctica se muestra el estándar de corrección del examinador/a principal. Mientras corrigen las respuestas de práctica, los examinadores/as pueden pedir orientación a su jefe/a de equipo si lo necesitan.

Una vez que los examinadores/as hayan terminado con las respuestas de práctica, deben tener la confianza de que comprenden y saben aplicar el estándar de corrección del examinador/a principal. Entonces, pueden comenzar el proceso de corrección de respuestas de aptitud.

Respuestas de aptitud

El propósito de las respuestas de aptitud es dar al equipo examinador la oportunidad de demostrar que sabe corregir según el estándar de corrección pertinente. Después de corregir adecuadamente las respuestas de aptitud, los examinadores/as podrán corregir en una convocatoria real.

Normalmente hay cinco respuestas de aptitud en un conjunto. Si un examinador/a no otorga las mismas puntuaciones que el examinador/a principal durante el proceso de corrección de respuestas de aptitud, se le proporcionan comentarios sobre qué puntuación debería haber otorgado y por qué. A continuación, al examinador/a se le ofrece una segunda oportunidad para demostrar que ahora comprende el estándar de corrección, para lo cual se utiliza un conjunto diferente de respuestas de aptitud.

Un buen conjunto de respuestas de aptitud debe incluir un rango de ejemplos de desempeño del alumnado, con puntuaciones de vayan de bajas a altas. Además, debe incluir ejemplos de respuestas que requieran que el examinador/a haya comprendido cualquier instrucción específica del esquema de calificación que le ayude a afrontar las dificultades previstas que el examinador/a principal y el equipo de examinadores/as supervisores consideren que pueden encontrarse durante la corrección. Dichas dificultades previstas se basarán en los hallazgos obtenidos en la revisión de los exámenes durante la reunión de estandarización. Las respuestas de aptitud no se seleccionan con la intención de “tender una trampa” a los examinadores/as, sino que representan los tipos de respuestas que probablemente encontrarán cuando corrijan.

En esta etapa, el personal examinador tiene dos oportunidades para demostrar que ha comprendido el estándar de corrección del examinador/a principal y que sabe aplicarlo al trabajo entregado para evaluación. Los examinadores/as deben corregir adecuadamente un conjunto de respuestas de aptitud.

Cada conjunto de respuestas de aptitud se compone de varios exámenes que están corregidos definitivamente. No se permite a los examinadores/as discutir con su jefe/a de equipo sobre dichos exámenes mientras corrigen un conjunto de respuestas de aptitud. Si la corrección de un examinador/a en

el primer conjunto de respuestas de aptitud está dentro de los márgenes de tolerancia, entonces pasa a la corrección en condiciones reales.

Sin embargo, si la corrección del examinador/a está fuera de los márgenes de tolerancia, el jefe/a de equipo le aporta comentarios sobre cómo ajustar su corrección al estándar del examinador/a principal. Tras reflexionar sobre estos comentarios, se le da una segunda oportunidad para que demuestre que ha comprendido el estándar de corrección.

Si la corrección de un examinador/a en el segundo conjunto de respuestas de aptitud está dentro de los márgenes de tolerancia, entonces se le permite comenzar a corregir en condiciones reales.

Corrección en condiciones reales y respuestas de control

Las respuestas de control forman parte del proceso de garantía de calidad. Su propósito supervisar el desempeño de los examinadores/as y demostrar que siguen corrigiendo de acuerdo con el estándar pertinente.

Con el paso del tiempo, es posible que las correcciones de un examinador/a se desvíen con respecto al estándar establecido por el examinador/a principal. Por ese motivo, la corrección del personal examinador se comprueba periódicamente mediante respuestas de control, que el examinador/a principal ya ha corregido definitivamente. Las respuestas de control tienen el mismo aspecto que el resto, por lo que el examinador/a no sabe que está corrigiendo una respuesta con fines de garantía de calidad. En general, los examinadores/as deben corregir una respuesta de control al azar por cada diez respuestas en condiciones reales.

Si el examinador/a concede la misma puntuación que el examinador/a principal en la respuesta de control, podrá continuar corrigiendo. Sin embargo, si concede una puntuación errónea, lo cual muestra que su corrección ya no sigue el estándar requerido, el IB interviene.

Inicialmente, los examinadores/as supervisores ofrecen comentarios y orientación para que el examinador/a tenga la oportunidad de ajustarse de nuevo al estándar de corrección. Si no es capaz de hacerlo pese a haber recibido apoyo adicional, no se le permite que corrija ninguna respuesta más.

Se establecen márgenes de tolerancia de corrección para cada componente, lo que permite cierto grado de variación con respecto al estándar definitivo. Para obtener más información, consulte la sección [“Márgenes de tolerancia”](#).

Indicadores de una estandarización satisfactoria

Un ejercicio de estandarización satisfactorio da como resultado:

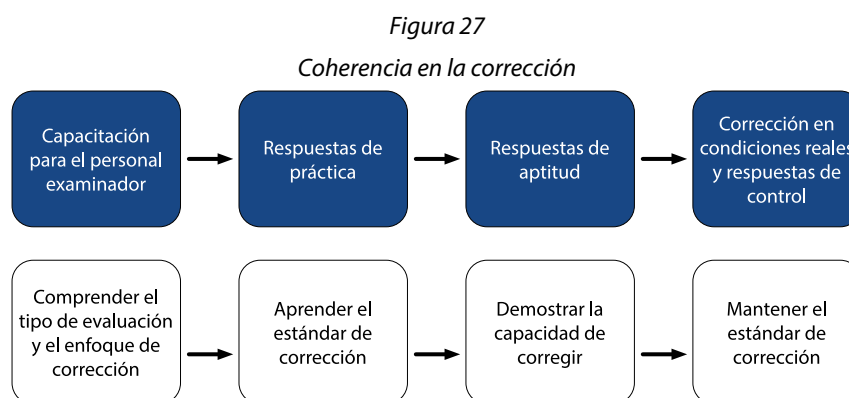
- Equipos de examinadores/as supervisores comprometidos que son capaces de ayudar al personal examinador con sus consultas sobre la corrección, y de garantizar la coherencia de esta
- Instrucciones claras y sin ambigüedades, en forma de esquema de calificación o de notas para la corrección, que expliquen claramente la fundamentación de las puntuaciones otorgadas
- Respuestas de práctica, de aptitud, y de control corregidas definitivamente para fines de capacitación y de garantía de calidad

Márgenes de tolerancia

El IB cree en hacer preguntas que evalúen lo que es importante, y no simplemente lo que es fácil de corregir. En los tipos de preguntas en las que la respuesta puede ser o correcta o incorrecta, es razonable esperar que el examinador/a conceda la misma puntuación que el examinador/a principal. En el caso de otros tipos de preguntas, sobre todo las que ponen a prueba la comprensión o el análisis, se requiere en cierta medida el juicio del examinador/a. En la práctica, esto puede significar que dos examinadores/as con la misma experiencia y comprensión de cómo se debe responder adecuadamente una pregunta no otorguen exactamente la misma puntuación por el mismo trabajo: puede haber una diferencia de 1 o 2 puntos.

La idea de utilizar márgenes de tolerancia es reflejar estas diferencias de opinión legítimas al revisar el desempeño de los examinadores/as. Por ejemplo, si el examinador/a principal concedió a un ensayo una puntuación de 46 y la pregunta tenía un margen de tolerancia de 2 puntos, entonces se considera que cualquier examinador/a que conceda una puntuación de entre 44 y 48 demuestra que corrige según el estándar establecido.

El IB determina la diferencia aceptable entre la puntuación definitiva (la que concede el examinador/a principal) y la puntuación que otorgan los examinadores/as en estos exámenes. Esta diferencia permitida se denomina *margen de tolerancia*. Cuando se revisan las preguntas que tienen distintas partes o que se corrigen con varios criterios, el IB supervisa la diferencia en la puntuación general y también las diferencias en cada parte o criterio. El examinador/a demostrará que corrige según el estándar si sus resultados se encuentran dentro de los márgenes de tolerancia en ambos aspectos.



Capacitación para el personal examinador

Antes de que comience una convocatoria de exámenes, se ofrece capacitación al personal examinador. El propósito de la capacitación es ayudarles a comprender el tipo de evaluación y el enfoque general para corregirla. La capacitación no contiene información específica sobre los exámenes de futuras convocatorias.

Respuestas complicadas o inusuales

Es posible que, debido a las respuestas del alumno/a, los examinadores/as se encuentren con exámenes que no sepan cómo corregir. En esos casos, deben pedir consejo a sus jefes/as de equipo. Si aun así siguen considerando que no son capaces de corregir ese examen de manera justa, pueden enviarlo a un examinador/a situado en un nivel superior en la jerarquía para que lo corrija. Los exámenes con respuestas especialmente problemáticas se pondrán en conocimiento del examinador/a principal, quien tomará una decisión definitiva sobre cómo corregirlos.

Asimismo, esta persona será la encargada de corregir los exámenes que tengan respuestas inusuales y los exámenes del alumnado que utilice cuestionarios modificados. El examinador/a principal deberá velar por que el estándar de corrección en estos exámenes sea equivalente al que se aplica al resto del alumnado.

Vínculos con los colegios

El IB parte del principio de que los examinadores/as no deben corregir el trabajo de su propio alumnado.

Para ello, se pide a todos los examinadores/as que informen al IB de cualquier conexión que puedan tener con estudiantes y colegios, a fin de que no se les envíe trabajos de estudiantes de colegios con los que exista un conflicto de intereses. Para obtener más información sobre este proceso, consulte la sección [“Conflictos de intereses”](#) de esta publicación.

Comentarios y anotaciones del personal examinador

El propósito de la evaluación sumativa del IB es medir el desempeño de cada estudiante. El personal examinador debe centrarse exclusivamente en corregir el trabajo del alumnado según el estándar requerido. El personal examinador solo debe incluir comentarios si esto le ayuda en su labor de corrección.

Para poder ofrecer comentarios formativos, ya sean para el alumnado o para el profesorado, el examinador/a debe determinar la puntuación correcta que corresponda al trabajo del alumno/a y, a continuación, intentar explicar cómo se podría mejorar dicho trabajo. El IB es consciente de que esta segunda tarea no es trivial, sino que constituye una de las características clave de una buena enseñanza. Sin embargo, esto requiere un tiempo y una reflexión que alejarían al examinador/a de su cometido principal, que es corregir según un estándar coherente. En términos sencillos, el IB quiere que el personal examinador realice una tarea —corregir— con la máxima calidad, y no dos tareas —corregir y proporcionar comentarios— con una calidad menor.

Un examinador/a solamente puede ofrecer su valoración a partir del trabajo que tiene a su disposición, mientras que el profesorado con experiencia puede utilizar una amplia variedad de información para decidir qué comentarios aportar a un alumno/a. Por tanto, los comentarios del examinador/a tendrán una perspectiva más limitada que los del profesor/a.

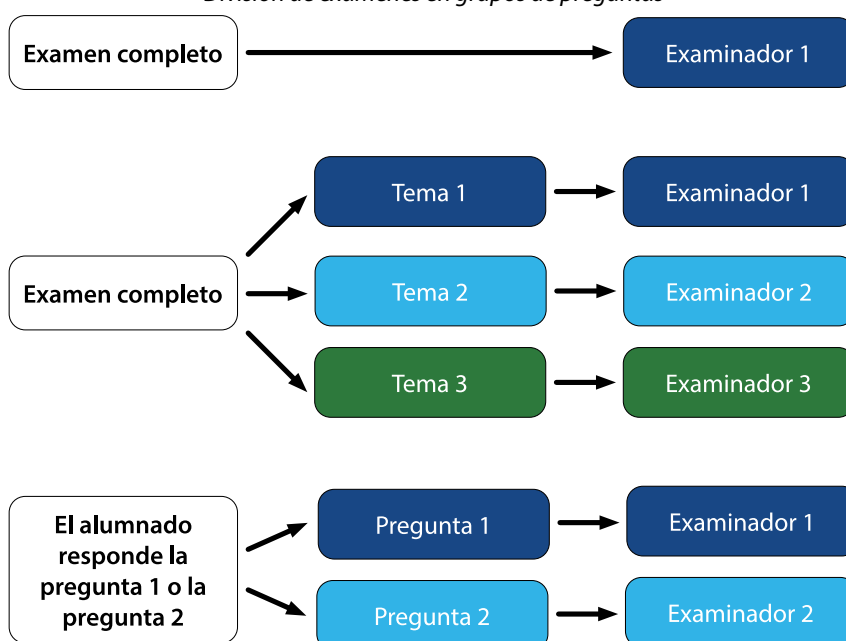
Por todos estos motivos, el IB deja muy claro al personal examinador que debe limitarse a corregir el trabajo del alumnado siguiendo el estándar establecido, y que no debe añadir comentarios para el alumno/a ni para el profesor/a.

Para facilitar el proceso de corrección, los examinadores/as deben indicar claramente en qué partes otorgan puntos y, en caso de ambigüedad, explicar su decisión mediante comentarios adecuados. Esto permite al IB comprobar los estándares de corrección y hace que el proceso de concesión de puntuaciones sea más transparente para los colegios.

Grupos de preguntas

Figura 28

División de exámenes en grupos de preguntas



La razón de ser de los grupos de preguntas es que para los examinadores/as es más fácil corregir la misma pregunta varias veces seguidas que corregir todas las preguntas del examen de un alumno/a y luego empezar de nuevo con el siguiente alumno/a.

Cada grupo de preguntas contará con sus propias respuestas de práctica, de aptitud y de control, y los examinadores/as deberán demostrar que son capaces de corregir según el estándar establecido para sus grupos de preguntas. Aunque pueda parecer complicado, esto significa que un examinador/a que no comprenda bien cómo se aplica el estándar a una pregunta concreta de un examen puede seguir corrigiendo las demás preguntas, en vez de no permitirle que corrija el examen entero por no cumplir el estándar establecido.

Totalización

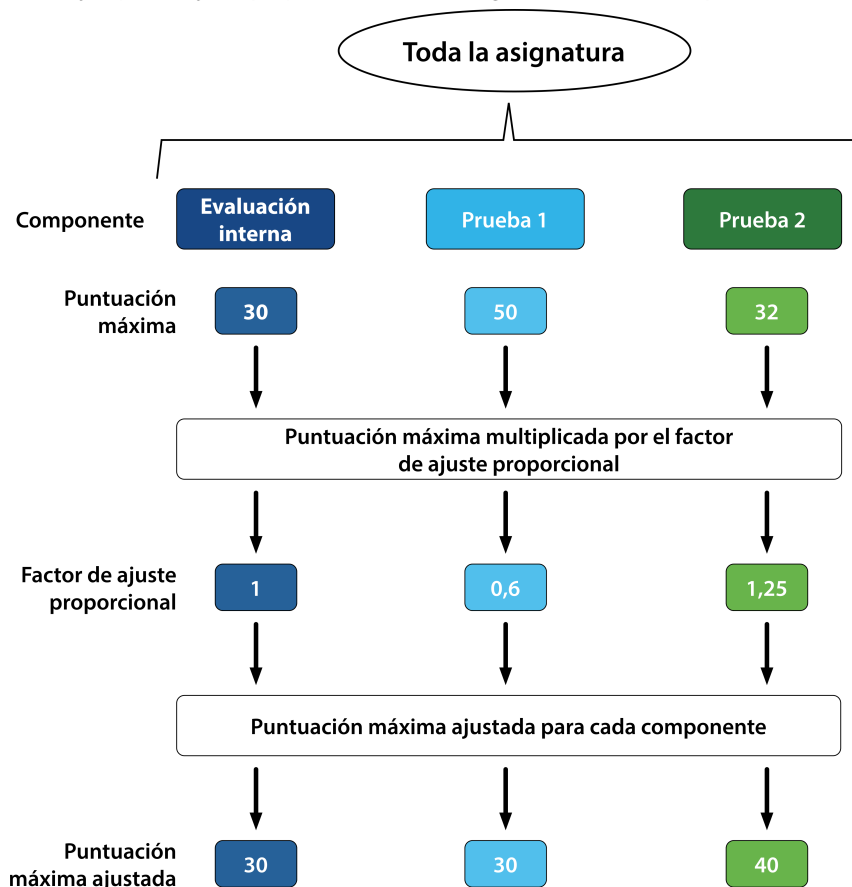
- La totalización es el proceso de combinar las puntuaciones obtenidas en los componentes para obtener un resultado total.
- Para que cada componente contribuya a la puntuación final en la medida que se desea (lo que se conoce como *porcentaje del total de la evaluación* o, en ocasiones, *ponderación*), puede que sea necesario ajustar proporcionalmente las puntuaciones de los componentes.
- El IB utiliza un *modelo de compensación* mediante el cual el alumnado puede compensar un mal desempeño en un componente o pregunta si obtiene un buen resultado en otro.
- La calificación final que un alumno/a obtiene en una asignatura se determina a partir de la totalización de las puntuaciones logradas en cada componente, y no de calificaciones por componente.

La totalización es el proceso de combinar las puntuaciones (y los límites) de los distintos componentes para obtener una puntuación final o un límite de calificación global. Para ello, puede que sea necesario ajustar proporcionalmente las puntuaciones globales (o los límites) de los componentes.

El objetivo del ajuste proporcional es conservar el valor porcentual que se quiere que tenga cada componente con respecto al total de la evaluación del curso. Esto implica multiplicar o dividir las puntuaciones de los componentes para que así representen la parte proporcional correcta del total global de la asignatura. Lo mismo sucede con los límites de calificación que se establecen para el componente, que se deberán determinar inicialmente a partir de las puntuaciones máximas para los componentes.

La puntuación máxima ajustada para cada componente es la misma que el porcentaje del total de la evaluación asignado al componente en la guía de la asignatura.

Figura 29
Ejemplo de ajuste proporcional en una asignatura con tres componentes



El porcentaje del total de la evaluación permite reflejar la importancia relativa que el IB otorga a los elementos que se evalúan. Por ejemplo, si un componente sirve sobre todo para evaluar la interpretación de datos o fuentes y representa un porcentaje del total de la evaluación del 30 %, eso implica que, en comparación con los demás objetivos específicos del curso, la interpretación representa aproximadamente el 30 % de lo que es importante. A menudo, varios componentes sirven para evaluar los mismos objetivos específicos, lo que hace que este cálculo resulte menos significativo.

Un aspecto secundario, aunque importante, del porcentaje del total de la evaluación es que permite al IB establecer en una evaluación la puntuación total que resulte adecuada para las tareas y para los criterios de corrección, en vez de intentar crear artificialmente un total global. A continuación, dichos totales se pueden totalizar para alcanzar una calificación final.

La totalización en la evaluación consiste en combinar las puntuaciones de diferentes componentes para determinar la calificación final de un alumno/a. Este enfoque permite flexibilidad para establecer puntuaciones en función de la dificultad y los criterios de la tarea y las intenciones del diseño de la evaluación, en lugar de ceñirse a un total fijo indicado por los porcentajes del total de la evaluación de los componentes. Esto implica que, aunque los porcentajes del total de la evaluación de los componentes se tienen en cuenta, los resultados reales pueden ser diferentes de los resultados previstos debido, por ejemplo, a la facilidad o dificultad comparativa imprevista de una evaluación o un elemento. El enfoque del IB no implica una equivalencia exacta entre puntuaciones individuales y porcentajes predeterminados del total de la evaluación; lo que más importa son los resultados generales. Este método, denominado *modelo de compensación*, permite al alumnado compensar un mal desempeño en un área si tiene un buen desempeño en otra, ya que se centra en la puntuación total en lugar de en las puntuaciones de cada componente. Por tanto, la calificación final se obtiene de la suma de todas las puntuaciones logradas en los componentes, y no del promedio de las calificaciones por componente. Esto implica que estudiantes con las mismas calificaciones por componente pueden recibir calificaciones finales diferentes.

Moderación

- La moderación es el proceso de comprobar los estándares de corrección del profesorado, no el proceso de revisar la corrección del trabajo del alumnado.
- Que una moderación esté bien hecha significa que cada estudiante recibiría la misma puntuación de evaluación interna aunque hubiera estudiado en otro colegio en la otra punta del mundo. El IB denomina a esto *estándar global*.
- Entre las pruebas que se utilizan en la moderación figuran no solo la calidad del trabajo del alumnado, sino también las explicaciones del profesorado acerca de por qué ha otorgado sus puntuaciones.

La moderación con toma dinámica de muestras se utiliza para la evaluación interna en el PD y el POP, y también para las carpetas electrónicas en el PAI.

Definición de la moderación

Muchas veces no es posible evaluar un aspecto concreto de un alumno/a mediante un examen formal y con limitación de tiempo. En dichos casos, la forma más válida de proceder es pedir al profesor/a que se encargue de esto mediante una evaluación interna. En la sección [“La función de la evaluación realizada en clase y de la evaluación interna”](#) se explica por qué este es un enfoque adecuado.

Aunque este enfoque produce resultados significativos, también implica el riesgo de que distintos profesores/as tengan distintas interpretaciones del estándar de corrección. En ese sentido, sería posible que dos profesores/as de colegios distintos otorgaran puntuaciones distintas al mismo trabajo. El IB sigue un proceso integral de capacitación y puesta a prueba de los examinadores/as, con la intención de que todos compartan una comprensión común del estándar que se requiere. Sin embargo, esto no es posible para todo el profesorado.

En lugar de ello, el IB confía en la capacidad del profesorado para corregir todos los trabajos de sus estudiantes con coherencia, por lo que, si es necesario, solo hay que aplicar un ajuste para que las puntuaciones que conceda estén en consonancia con el estándar global. Para hacer esto, el IB pide una muestra de las correcciones del profesorado y la compara con el estándar del examinador/a principal. A partir de los datos que se desprenden de esta comparación, cuando es necesario, el IB calcula una fórmula matemática que permite ajustar las puntuaciones que ha otorgado cada docente. En función de una comparación estadística entre los dos grupos de puntuaciones, si es necesario, se realiza un ajuste a las puntuaciones que el profesor/a otorgó en ese componente a todo el alumnado del colegio. Si las puntuaciones que otorga el profesor/a son sistemáticamente demasiado bajas o demasiado altas, se realizará el mismo ajuste en todas las puntuaciones que otorgue. Sin embargo, si las puntuaciones que otorga el profesor/a son demasiado bajas o demasiado altas únicamente en la parte superior o en la parte inferior del rango de puntuaciones, es posible que el ajuste varíe en todo el rango de puntuaciones del profesor/a.

Información importante que hay que recordar

- El objetivo de la moderación es comprobar con qué precisión y coherencia ha aplicado el profesor/a los criterios de evaluación al corregir el trabajo del alumnado.
- Como resultado de la moderación, las puntuaciones de un colegio pueden subir o bajar. También pueden quedarse igual.
- Si se aplica un factor de moderación, esto no implica que la corrección del profesor/a tuviera poca calidad, simplemente que no ha sido coherente con el estándar global.

Por razones prácticas, el IB modera a los colegios y no a docentes individuales. Por ese motivo, es muy importante que todo el profesorado de un colegio se asegure de corregir según el mismo estándar. El factor de moderación se aplica a todas las puntuaciones del profesorado, por lo que la muestra debe ser una representación fiel de la corrección realizada en el colegio para garantizar que cualquier ajuste sea adecuado.

Aspectos que debe tener en cuenta el personal de coordinación al revisar las puntuaciones de la evaluación interna que se moderarán:

- En el colegio, ¿todo el profesorado (encargado de otorgar las puntuaciones de la evaluación interna) corrige según el mismo estándar?
- ¿El profesorado corrige de manera coherente a todo el alumnado según el mismo estándar?
- ¿El profesorado explica claramente por qué concede una puntuación concreta?

Para obtener más información sobre la aplicación precisa de los criterios de evaluación, consulte la sección [“Comentarios para los colegios”](#) de esta publicación, así como las publicaciones *Procedimientos de evaluación* vigentes del PAI, el PD y el POP.

Para obtener más información sobre los aspectos técnicos de cómo se calcula el factor de moderación, consulte el apéndice [“La moderación de la evaluación interna en mayor detalle”](#) de esta publicación.

Moderación y expectativas del profesorado

Los trabajos entregados para evaluación los evalúa el personal docente internamente y los modera el IB externamente.

El profesorado debe familiarizarse con la guía de la asignatura pertinente, en la que se proporciona información detallada sobre los criterios de la evaluación interna. Cada criterio de evaluación cuenta con descriptores que describen un nivel de logro específico y equivalen a un determinado rango de puntos. Los descriptores de nivel se centran en aspectos positivos, aunque, en los niveles más bajos, la descripción puede mencionar la falta de logros.

El profesorado debe valorar el trabajo de evaluación interna con relación a los criterios, utilizando los descriptores de nivel y las aclaraciones. Los criterios deben aplicarse utilizando el descriptor más adecuado para el trabajo. Cuando un trabajo muestre niveles de logro distintos para los diferentes aspectos de un criterio, la puntuación asignada debe ser aquella que refleje más justamente el logro general de los aspectos del criterio. No es necesario cumplir todos los aspectos de un descriptor de nivel para obtener dicha puntuación. Los descriptores de nivel más altos no implican que el trabajo del alumno/a es perfecto.

En los casos en que un descriptor de nivel comprenda dos o más puntuaciones, el equipo docente debe conceder la puntuación más alta si el trabajo se corresponde en gran medida con las cualidades descritas (el trabajo puede estar cerca de alcanzar las puntuaciones del descriptor de nivel superior). Se deben conceder las puntuaciones más bajas si el trabajo demuestra en menor medida las cualidades descritas (el trabajo puede estar cerca de alcanzar las puntuaciones del descriptor de nivel inferior).

Solamente deben utilizarse números enteros. No se aceptan puntuaciones parciales, como fracciones o decimales.

Cada criterio debe considerarse de forma independiente. Un alumno/a que alcance un nivel de logro alto en un criterio no necesariamente alcanzará niveles altos en los demás criterios. Igualmente, un alumno/a que alcance un nivel de logro bajo en un criterio no necesariamente alcanzará niveles bajos en los demás criterios. El profesorado no debe suponer que la evaluación general de sus estudiantes debe dar como resultado una distribución determinada de puntuaciones.

En algunas asignaturas, los descriptores de nivel pueden hacer referencia a términos de instrucción específicos. Dichos términos de instrucción deberán interpretarse según lo dispuesto en la sección pertinente de la guía.

El profesorado debe ser consciente de que cuando se aplica un factor de moderación a los trabajos que ha evaluado internamente, esto no significa que haya seguido de manera errónea el procedimiento de corrección. El ajuste se realiza simplemente para que su corrección esté en consonancia con el estándar global establecido. El informe general de cada asignatura, que se publica después de cada convocatoria de exámenes, suele contener información útil para el personal docente sobre el desempeño en la evaluación interna. Además, en el Centro de Recursos para los Programas están disponibles materiales de ayuda al profesor.

A todos los trabajos que un colegio entrega para evaluación se les aplica el mismo factor de moderación. Por tanto, si varios profesores/as comparten la responsabilidad de la evaluación interna en la misma asignatura dentro de un colegio, es fundamental que se realice una estandarización interna para que el estándar de corrección dentro del colegio sea lo más coherente posible. Esto puede implicar discutir los criterios de evaluación y revisar los trabajos entregados para evaluación antes de realizar cualquier corrección.

El profesorado y el alumnado deben comprender los conceptos relacionados con la integridad académica, especialmente los de *autoría original* y *propiedad intelectual*. Los trabajos del alumnado deben ser originales en su totalidad y estar preparados de acuerdo con los requisitos que figuran en la guía de cada asignatura.

A todo el alumnado debe quedarle clara la diferencia entre colaboración y colusión. Para obtener más información, consulte la sección “Fomento de la mentalidad ética”.

El profesorado que desee acceder a un resumen de esta información puede imprimir el póster “Corrección de la evaluación interna: expectativas del profesorado en el proceso de moderación”, disponible en el apéndice “Recursos para imprimir” de esta publicación.

Selección de trabajos del alumnado

- Todo el alumnado debe recibir una puntuación justa, por lo que el trabajo de cualquier alumno/a debe poder ilustrar el estándar del profesor/a y servir de ejemplo para la moderación.
- Es posible que distintos profesores/as tengan ideas diferentes sobre lo que constituye un muy buen trabajo o un muy mal trabajo, por lo que es importante que la muestra de su colegio cubra todo el rango de puntuaciones otorgadas.
- En aras de la transparencia y la equidad entre todos los colegios, el IB debe poder consultar pruebas que muestren todo el trabajo que conforme la puntuación final de un alumno/a.

Con el objetivo de garantizar la transparencia y evitar cualquier impresión de conducta impropia, es importante que el IB, y no el colegio, sea el encargado de identificar el trabajo que se va a tomar como muestra para la moderación. El IB se basa en principios generales para garantizar que el factor de moderación sea lo más fiable posible y, dentro de las limitaciones de estos principios, que el alumnado se seleccione de manera aleatoria.

El primer principio es que el IB debe utilizar la muestra más reducida posible para obtener un factor de moderación fiable para el colegio en cuestión. De esta manera se minimizan la carga de trabajo para los colegios y los costos para el IB que, en realidad, luego repercutirían en los colegios en forma de tasas de examen. Sin embargo, es razonable que la cantidad de trabajos de estudiantes que se pida incluir en la muestra sea diferente en función del colegio si el IB necesita más ejemplos de la corrección del profesorado para determinar un factor de moderación sólido.

El segundo principio es que el IB debe estar seguro de que el factor de moderación es justo para todo el alumnado en todo el rango de puntuaciones. La experiencia demuestra que el profesorado puede tener distintas expectativas en varios niveles del rango de puntuaciones, y que un profesor/a que sea más generoso que el estándar global con los trabajos de mala calidad puede también ser más severo que el estándar global con los trabajos de buena calidad. Por este motivo, la muestra de la evaluación interna se selecciona cuidadosamente para garantizar que el rango de puntuaciones del colegio esté representado adecuadamente.

Para la muestra que se somete a moderación, el IB no suele seleccionar a estudiantes que hayan logrado la máxima puntuación. Esto permite que el alumnado que esté dentro del rango de puntuaciones más elevado tenga la posibilidad de que sus puntuaciones suban a raíz de la moderación al alza si el profesor/a ha sido demasiado severo en la corrección. Asimismo, normalmente el IB no selecciona trabajos de estudiantes que han obtenido una puntuación de 0.

Situaciones irregulares

El objetivo de la moderación es comprobar que el profesorado corrige de acuerdo con el estándar global, aun si un trabajo es particularmente difícil de corregir. En esos casos, es posible que el profesor/a tenga que aportar una justificación más detallada de por qué ha concedido esa puntuación. Por ejemplo, si un alumno/a ha recibido una ayuda adicional por parte del profesor/a para completar su trabajo, el profesor/a debe explicarlo en sus comentarios sobre la corrección, y el examinador/a debe tener esto en cuenta al revisar la corrección de ese trabajo.

Casos en los que no es posible encontrar un factor de moderación

En algunos casos, tal vez no sea posible calcular un factor de moderación con el trabajo de muestra entregado. Eso sucede si la diferencia entre la puntuación del examinador/a y la del profesor/a resulta incoherente, o si la puntuación del examinador/a indica que la puntuación del profesor/a ha sido demasiado generosa o demasiado severa en comparación con el estándar global. En dichos casos, el IB pedirá al colegio que envíe más trabajos para que se pueda aplicar un factor de moderación justo. Por este motivo, todos los trabajos del alumnado deben estar disponibles hasta que finalice la convocatoria de exámenes pertinente. En casos excepcionales, si no se puede aplicar un factor de moderación justo, se otorgará la puntuación que determine el examinador/a.

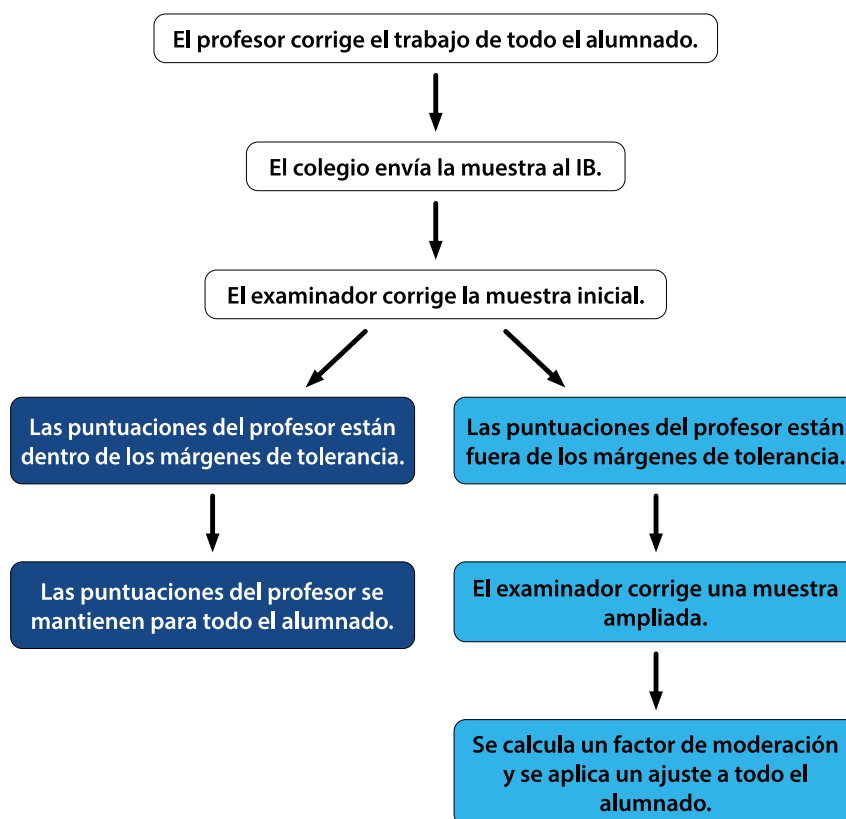
Toma dinámica de muestras

- La moderación con toma dinámica de muestras significa que el IB acepta las puntuaciones de un profesor/a si se encuentran dentro de los márgenes de tolerancia.
- Si alguna de las puntuaciones que otorga el profesor/a se encuentra fuera de los márgenes de tolerancia, el IB aplica un factor de moderación.
- El trabajo de todo el personal examinador que comprueba las correcciones de la evaluación interna se supervisa con el fin de garantizar su calidad.

El principio de la moderación con toma dinámica de muestras implica que si las puntuaciones que otorga el profesor/a en la muestra se encuentran dentro de los márgenes de tolerancia, demostrando así que comprende el estándar global, se considera que todas sus puntuaciones respetan el estándar requerido. Si, por el contrario, las puntuaciones están fuera de los márgenes de tolerancia, entonces se calcula y se aplica un factor de moderación. Esto se explica en la figura 30.

Figura 30

Moderación con toma dinámica de muestras



Con el fin de que la moderación sea justa, los examinadores/as tienen que comprender el estándar global que establezca el examinador/a principal y deben realizar la moderación según este estándar. Los examinadores/as encargados de la moderación reciben capacitación sobre el estándar global, y después se comprueba que realicen la moderación según este estándar. Tal como se describe en la sección “Estandarización”, el examinador/a principal prepara tres tipos de trabajos de evaluación interna corregidos definitivamente:

- **Respuestas de práctica**, para explicar el estándar global
- **Respuestas de aptitud**, para comprobar que los examinadores/as comprenden el estándar global
- **Respuestas de control**, para comprobar que los examinadores mantienen el estándar global

El mismo examinador/a revisará todos los trabajos de evaluación interna de un mismo colegio. Sin embargo, los examinadores/as reciben trabajos de distintos colegios sin ningún orden concreto. De esta manera, se evita que los examinadores/as se formen una opinión sobre si un profesor/a puntúa demasiado alto o demasiado bajo al ver el trabajo del primer alumno/a, y que busquen ese patrón en el resto de la muestra del colegio. Además, este método permite incluir respuestas de control sin que resulte obvio que lo son. Una vez que se haya determinado si es necesario aplicar un factor de moderación, se puede pedir a los examinadores/as que proporcionen comentarios breves acerca de la corrección del colegio.

El modelo de calidad es más complejo con la toma dinámica de muestras que con otros métodos de corrección. El IB también debe asegurarse de que el personal examinador sea capaz de determinar en qué casos la corrección del profesorado es demasiado severa, demasiado generosa, o correcta. En la figura 31 se muestra el rango de trabajos de evaluación interna necesario para el modelo de calidad a fin de cubrir todas las situaciones.

Figura 31

Rango de trabajos de evaluación interna necesario para el modelo de calidad

	Situación	
La puntuación del examinador está en la parte superior del rango de puntuaciones	1	La puntuación del profesor es demasiado severa
	2	La puntuación del profesor es acorde con el estándar global
	3	La puntuación del profesor es demasiado generosa
La puntuación del examinador está en la parte media del rango de puntuaciones	4	La puntuación del profesor es demasiado severa
	5	La puntuación del profesor es acorde con el estándar global
	6	La puntuación del profesor es demasiado generosa
La puntuación del examinador está en la parte inferior del rango de puntuaciones	7	La puntuación del profesor es demasiado severa
	8	La puntuación del profesor es acorde con el estándar global
	9	La puntuación del profesor es demasiado generosa

Para obtener más información sobre cómo modera el IB, consulte el video [“Los dos métodos para la corrección: cómo realizan la corrección los examinadores y los profesores”](#).

Concesión de calificaciones finales y totalización

- Las calificaciones finales deben tener el mismo significado, independientemente de la convocatoria en la que el alumno/a realice el examen.
- En el proceso de concesión de calificaciones se decide cómo convertir las puntuaciones en calificaciones finales para garantizar que este sea el caso.
- Los límites de calificación se definen utilizando varias pruebas, como los resultados del grupo de estudiantes y la valoración, por parte de especialistas, del trabajo entregado para evaluación.
- Lo más importante es la calificación final global que recibe el alumno/a, y no los componentes individuales.

Es importante recordar que la puntuación y la calificación son conceptos distintos. Para comprender mejor cuál es la diferencia, consulte la sección [“Diferencia entre puntuaciones y calificaciones”](#). En el proceso de concesión de calificaciones se decide qué correspondencia hay entre las puntuaciones y las calificaciones finales en un caso concreto.

Para tomar esta decisión, el examinador/a principal, el examinador/a jefe, los examinadores/as supervisores y el IB mantienen una serie de discusiones durante varios días. En este proceso se tienen en cuenta diversas pruebas (que se describen más detalladamente en la sección [“Pruebas utilizadas en el proceso de concesión de calificaciones finales”](#)) para llegar a una conclusión que sea justa para el alumnado de ese año, pero también para el alumnado que ya cursó la asignatura en el pasado. Finalmente, el examinador/a jefe hace una serie de recomendaciones al IB sobre cuáles deberían ser los resultados para esa convocatoria de exámenes.

El objetivo más destacado de estas discusiones es establecer los límites de calificación, donde se define la cantidad mínima de puntos que un alumno/a necesita para conseguir cada calificación final. Como no es práctico emitir un juicio detallado para cada calificación final de cada asignatura, el IB pide a los examinadores/as que recomienden varios *límites de calificación estimativos* y calcula el resto de los límites aritméticamente. Para obtener más información, consulte la sección [“Límites de calificación estimativos e interpolados”](#).

Los límites de calificación no serán necesariamente los mismos en cada convocatoria de exámenes. Esto se debe a que las preguntas que se plantean al alumnado son diferentes en cada iteración de un examen, y los límites de calificación deben modificarse para reflejar tareas más fáciles o más difíciles. El IB hace todo lo posible por garantizar que los exámenes tengan la misma dificultad cada año. Sin embargo, dada la importancia de lo que hay en juego, no se realizan exámenes de práctica, debido al riesgo de que se divulguen las preguntas.

Al decidir dónde situar los límites de calificación, los examinadores/as y el IB tienen en cuenta los aspectos que se muestran en la figura 32.

Figura 32

Aspectos que se tienen en cuenta durante el proceso de concesión de calificaciones finales

Grupo de estudiantes	<ul style="list-style-type: none"> • ¿Cómo es el grupo de estudiantes actual que realiza el examen en comparación con grupos de años anteriores?
Examen	<ul style="list-style-type: none"> • ¿En qué medida el examen ha cumplido las expectativas que tenía el autor? • ¿Qué opinó el profesorado sobre la dificultad del examen?
Exámenes escritos del alumnado	<ul style="list-style-type: none"> • ¿Cómo es el trabajo del alumnado en comparación con el de años anteriores y con los descriptores de calificaciones finales?
Resultados	<ul style="list-style-type: none"> • ¿Cómo son los resultados del alumnado en comparación con los de años anteriores?
Equilibrio	<ul style="list-style-type: none"> • ¿Todas las pruebas disponibles llevan a una sola conclusión, o son contradictorias?

En muchas ocasiones, el IB cuenta con varios componentes de evaluación que se combinan para alcanzar una calificación general. En la sección “Totalización” de esta publicación se ofrece más información sobre este tema. El aspecto fundamental que el equipo examinador debe recordar durante la concesión de calificaciones finales es que lo más importante es el **resultado total**. Si es necesario, los límites de calificación de algunos componentes pueden ser menos perfectos para lograr un resultado global justo.

Formalmente, el propósito del proceso de concesión de calificaciones finales es:

- Determinar, a partir de los trabajos entregados para evaluación, en qué puntos hay un cambio en la calidad del trabajo que justifique pasar a otro descriptor de calificaciones finales
- Determinar, para cada componente, las puntuaciones que conformarán los límites de calificación
- Asegurarse de que el conjunto de los límites de calificación de cada asignatura da como resultado una concesión de calificaciones justa

El proceso de concesión de calificaciones finales se ha realizado de manera satisfactoria si:

- La valoración de los exámenes con las respuestas del alumnado y las pruebas obtenidas de los resultados coinciden ampliamente, teniendo en cuenta la información sobre el grupo de estudiantes
- Se pueden explicar las variaciones significativas que haya en el desempeño de los colegios
- El examinador/a jefe y el director/a en jefe de evaluación del IB consideran que se han mantenido los estándares de evaluación

Límites de calificación estimativos e interpolados

Los **límites de calificación estimativos** son los que recomienda el examinador/a jefe a partir de las discusiones mantenidas durante el proceso de concesión de calificaciones. En el PAI, el PD y el POP, los límites de calificación estimativos son 2-3, 3-4 y 6-7. El resto de límites de calificación (1-2, 4-5 y 5-6) se calculan aritméticamente basándose en los límites de calificación estimativos, por lo que se denominan **límites de calificación interpolados**.

Si hay un cambio significativo en la proporción del alumnado que recibe una calificación de 5 o de 1, hará falta discutir los resultados teniendo en cuenta el contexto del grupo de estudiantes. Esto puede conducir a que se reconsideren los límites estimativos o, en circunstancias excepcionales, a revisar los trabajos entregados para evaluación que se encuentren en esos límites.

Pruebas utilizadas en el proceso de concesión de calificaciones finales

La concesión de calificaciones finales es un proceso que se basa en pruebas que se obtienen de distintas fuentes de información, entre ellas:

- Los comentarios del profesorado sobre la evaluación
- El juicio experto del personal examinador sobre los ejemplos de trabajos del alumnado
- Un análisis de información estadística que compare los resultados de la convocatoria de exámenes en curso con los de años anteriores

Ningún tipo de prueba es más importante que otro: todas se deben tener en cuenta por igual para alcanzar conclusiones.

Consideración del grupo de estudiantes del año en curso

La primera tarea del proceso de concesión de calificaciones finales es plantearse en qué medida el grupo de estudiantes que realiza la evaluación se parece al de años anteriores. Si muchos de los colegios que realizan la evaluación el año en curso son los mismos que en años anteriores, es probable que las diferencias que haya en el desempeño se deban a la dificultad de los exámenes. Si este no es el caso (por ejemplo, si hay muchos colegios que inscriben a estudiantes en una asignatura por primera vez, o si hay una proporción elevada de estudiantes que vuelven a realizar la evaluación), entonces el IB podría decidir que es necesario que las calificaciones finales tengan en cuenta las diferencias en el desempeño.

Entre los factores que se tendrán en cuenta al comparar distintos grupos de estudiantes que realizan la evaluación, se pueden indicar:

- Cambios en el número de estudiantes que realizan las evaluaciones
- Cambios en la proporción de estudiantes que realizan las evaluaciones en español, francés e inglés (y en otras lenguas según corresponda)
- El número de nuevos colegios y la cantidad de estudiantes que haya en esos colegios
- Cambios en las opciones que elija el alumnado

Comentarios y opiniones sobre la evaluación

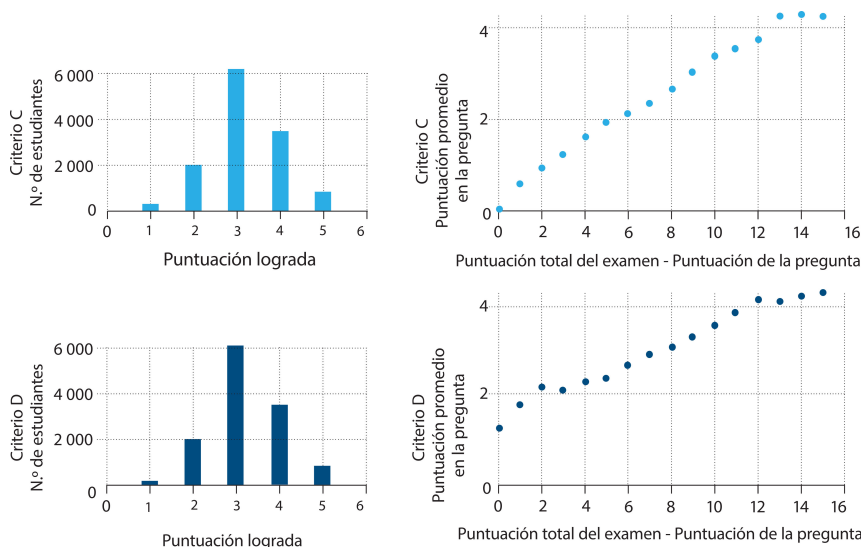
La siguiente tarea es plantearse cómo se desarrolló la evaluación en comparación con las expectativas que había. Así, si una pregunta en particular resulta ser mucho más complicada de lo que se esperaba, o si en las preguntas en las que se esperaba poder diferenciar entre el alumnado merecedor de una calificación de 6 y el merecedor de una calificación de 7 no se logra este propósito, entonces el equipo de concesión de calificaciones tendrá esto en cuenta cuando determine los límites de calificación.

El equipo de concesión de calificaciones se basará en los comentarios del profesorado sobre los exámenes y en las estadísticas sobre el desempeño en preguntas o elementos concretos (véase la figura 33). El equipo también cuenta con informes resumidos de distintos examinadores/as, y puede apoyarse en su propia experiencia de corregir trabajos de estudiantes durante la convocatoria de exámenes.

Figura 33

Ejemplo de las estadísticas que el personal examinador tiene a su disposición acerca de las preguntas

Resumen de las estadísticas sobre las preguntas							Correlación entre la pregunta y el resto del componente	
Pregunta	Preguntas que se han intentado contestar	Total de estudiantes para el examen	Proporción de estudiantes que han intentado contestar	Puntuación promedio en la pregunta	Puntuación más alta conseguida en la pregunta	Desv. estándar en la puntuación de la pregunta	Pregunta	Coefficiente de correlación
01 Criterio A	12.071	42.042	28,71 %	2,7	5	0,95	01 Criterio A	0,8044
01 Criterio B	12.071	42.042	28,71 %	2,6	5	0,94	01 Criterio B	0,8018
01 Criterio C	12.071	42.042	28,71 %	2,8	5	0,84	01 Criterio C	0,7983
01 Criterio D	12.071	42.042	28,71 %	3,2	5	0,79	01 Criterio D	0,7253
02 Criterio A	29.848	42.042	71,00 %	2,8	5	0,98	02 Criterio A	0,7954
02 Criterio B	29.848	42.042	71,00 %	2,7	5	0,93	02 Criterio B	0,7944
02 Criterio C	29.848	42.042	71,00 %	2,9	5	0,83	02 Criterio C	0,7865
02 Criterio D	29.848	42.042	71,00 %	3,2	5	0,79	02 Criterio D	0,7320



Búsqueda de pruebas en los exámenes

A continuación, los examinadores/as supervisores comparan los descriptores de cada calificación final con los exámenes que tienen puntuaciones cercanas a los límites de calificación previstos. Durante esta fase del proceso de concesión de calificaciones, los examinadores/as no deben centrarse en las puntuaciones otorgadas, sino en la naturaleza de las respuestas del alumnado y en qué medida encajan con los descriptores de calificaciones finales.

Antes de empezar a revisar los exámenes con las respuestas del alumnado, resulta útil para los examinadores/as mirar algunos ejemplos de años anteriores con el fin de recordar las expectativas correspondientes a las distintas calificaciones. Dichos ejemplos deben estar disponibles durante el proceso de concesión de calificaciones finales.

Antes de la reunión de concesión de calificaciones, el equipo de examinadores/as supervisores, y en concreto los examinadores/a principales, presentarán unos límites provisionales de calificación para cada componente, en los que indicarán en qué puntuaciones creen que deberían establecerse los límites, basándose en su experiencia de años anteriores acerca de los estándares previstos. Después de conocer estos límites provisionales, el consenso alcanzado sobre cómo ha funcionado cada examen, y la distribución general de puntuaciones, la persona responsable de la asignatura del IB dispone de la información y las pruebas que necesita para sugerir el rango de puntuaciones que estará representado en una muestra para la concesión de calificaciones finales.

Cada examinador/a supervisor debe revisar los exámenes con las respuestas del alumnado que se hayan seleccionado, y determinar qué descriptor de calificaciones finales refleja mejor la calidad del trabajo. Esta tarea resulta complicada, ya que incluso el alumnado que mantenga en su examen un nivel relativamente homogéneo mostrará a menudo en sus respuestas cualidades correspondientes a calificaciones muy

variadas. Es aceptable, y a menudo útil, indicar si un examen con las respuestas del alumno/a corresponde plenamente al descriptor de una calificación final o si está cerca de la siguiente calificación superior (esto se suele indicar de la siguiente manera: "7-" o "6+", por ejemplo).

Para llevar a cabo su cometido, es esencial que los examinadores/as eviten todo tipo de ideas preconcebidas y sesgos. Por tanto, es importante que no compartan sus puntos de vista hasta que todos hayan registrado sus resultados de manera independiente. Del mismo modo, es mejor no informar a los examinadores/as (excepto al examinador/a jefe) acerca de los resultados del análisis estadístico hasta que hayan finalizado esta etapa, excepto si se trata de información pertinente sobre el desempeño en preguntas o elementos concretos en todo el grupo de estudiantes.

Determinar qué calificación final corresponde a un examen con las respuestas del alumno/a es un ejercicio muy subjetivo, y es probable que haya variaciones entre distintos examinadores/as. Es más, sería razonable que un alumno/a que haya demostrado una buena comprensión, aunque haya cometido varios errores, reciba una calificación más alta que un alumno/a que haya destacado en la tarea más fácil (y haya obtenido por ello muchos puntos) pero que no haya demostrado la misma profundidad de comprensión, aunque este último alumno/a tuviera una puntuación ligeramente superior. Asimismo, es importante recordar que en esta revisión de exámenes solo se utiliza un número relativamente reducido de ejemplos de trabajos de estudiantes.


Hay investigaciones que indican que los examinadores/as tienen mucha habilidad para identificar en qué momento los exámenes muestran un nivel de calidad diferente al de otros que ya han revisado (véase la sección "Enfoques de la corrección" de esta publicación). El IB pide a los examinadores/as que empiecen por la puntuación más alta de la muestra y sigan en orden descendente hasta que consideren que han llegado a un punto en el que no encuentran de manera constante pruebas de las cualidades correspondientes a la calificación más alta. Entonces deben empezar por la puntuación más baja de la muestra y seguir en orden ascendente hasta que encuentren de manera constante pruebas de las cualidades correspondientes a la calificación más alta. Si es necesario, se puede aumentar el rango de puntuaciones de la muestra.

Una vez recopiladas todas las decisiones de los examinadores/as supervisores para conceder las calificaciones finales, dichas decisiones deben indicar el rango de puntuaciones en el que situar un límite de calificación. Esto se denomina *zona de incertidumbre*, y va desde la puntuación más baja hasta la más alta para las que hay pruebas razonables para colocar el límite de calificación. No existe una definición precisa de cómo establecer una zona de incertidumbre, así que los examinadores/as supervisores deben llegar a un acuerdo al respecto en función de sus decisiones individuales.

Figura 34

Ejemplo de resultados de la revisión de exámenes

Número de examen	Puntuación	Examinador 1 (examinador principal)	Examinador 2	Examinador 3	Examinador 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-



La única excepción a este proceso se da con los exámenes con preguntas de opción múltiple. La experiencia demuestra que emitir juicios sobre los límites de calificación en función de la calidad del trabajo entregado para evaluación es muy difícil en aquellos exámenes que solamente incluyen preguntas de opción múltiple. Esto se debe a que las respuestas no muestran qué ha hecho realmente el alumnado y, por

tanto, se carece de base suficiente para tomar una decisión. En ese tipo de exámenes, los límites de calificación se calculan de modo que el porcentaje de estudiantes de cada calificación final sea el más cercano posible al obtenido a partir de una estimación en los exámenes con los que guarden una relación más estrecha.

Análisis de las estadísticas sobre los resultados

Puede argumentarse que, en un sistema basado en criterios que depende de la opinión profesional de los examinadores/as, los examinadores/as supervisores deberían poder establecer los límites de calificación simplemente teniendo en cuenta las preguntas del examen y las respuestas que se esperan del alumnado. Sin embargo, en la realidad es muy difícil que dichas opiniones puedan formularse antes de conocer cómo ha respondido realmente el alumnado. Cresswell (2000) llega a la conclusión de que quienes corrigen suelen identificar acertadamente si los exámenes resultan más fáciles o más difíciles de una convocatoria a otra, pero habitualmente no logran estimar bien en qué medida eran más fáciles o más difíciles.

El propósito de la reunión de concesión de calificaciones finales es mantener un significado o un estándar coherentes de la calificación, por lo que resulta interesante reflexionar sobre una de las definiciones de *equiparabilidad* que proporciona Cresswell.

Se puede considerar que dos exámenes tienen estándares equiparables si dos grupos de estudiantes con las mismas capacidades y los mismos logros alcanzados, que han acudido a colegios parecidos con políticas de admisión idénticas, que han tenido docentes igual de competentes, y que han mostrado el mismo nivel de motivación, obtienen unas calificaciones finales que siguen la misma distribución después de haber cursado sus respectivos programas de estudios y de haber realizado los exámenes.

(Cresswell, 1996, pp. 57-84)

Para facilitar su labor de concesión de calificaciones finales, se proporciona a los examinadores/as supervisores los **límites recomendados estadísticamente**. Estos son límites de calificación que ofrecerían exactamente el mismo porcentaje acumulado (la proporción total de estudiantes que reciben la calificación dada o superior) que el año anterior hasta calificación dada.

Además, a los examinadores/as que participan en el proceso de concesión de calificaciones finales se les proporciona información sobre la puntuación media que obtiene el alumnado, así como histogramas sobre la distribución real de las puntuaciones. Aunque esta información se tiene en cuenta en los límites recomendados estadísticamente, suele ser útil analizarla con mayor detalle.

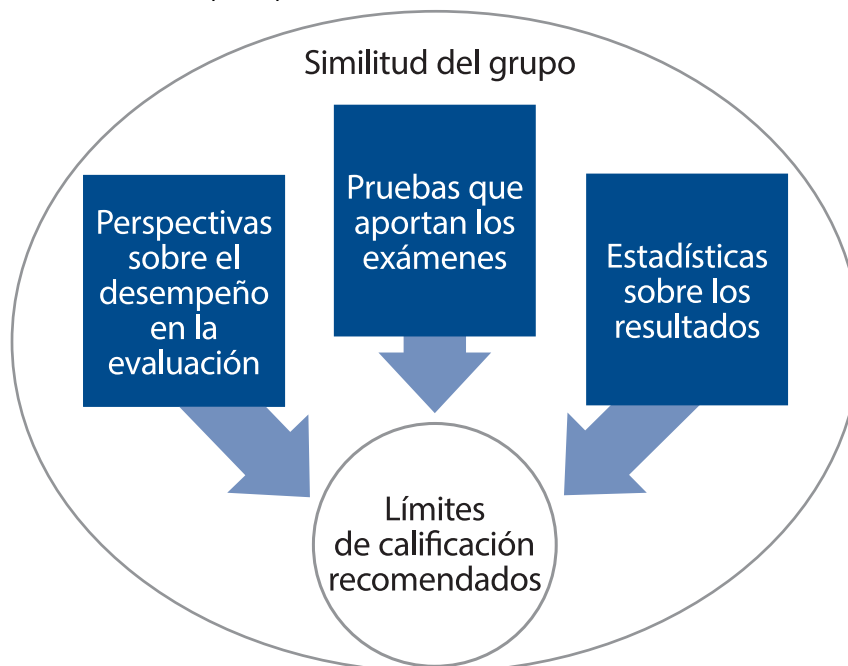
Es importante recordar que el grupo de estudiantes que realizan los exámenes cambia en cada convocatoria. Por ese motivo, es probable que los postulados de la definición de *equiparabilidad* que se da en la cita anterior no se puedan cumplir del todo. Así sucederá especialmente si se trata dos grupos de estudiantes muy distintos entre sí. De la misma manera, cuando hay una gran cantidad de estudiantes con experiencias educativas similares (por ejemplo, que provengan de Colegios del Mundo del IB), es más probable que las diferencias considerables en los resultados se deban a que dichos alumnos/as hayan respondido grupos de preguntas distintos, y no a variaciones en su desempeño.

Al considerar la importancia de los límites recomendados estadísticamente se debe tener en cuenta el tamaño de los grupos de estudiantes. El IB no tiene reglas formales en este aspecto, ya que hay otros factores que también son importantes. En líneas generales, si solamente hay varias decenas de estudiantes, se puede esperar que haya una variación considerable en el resultado global, pero es relativamente poco probable que esto suceda si hay miles de estudiantes.

Cómo encontrar el equilibrio entre las pruebas

Ningún tipo de prueba es más importante que otro. La tarea del equipo de concesión de calificaciones es tenerlas todas en cuenta para poder hacer una recomendación.

Figura 35
Pruebas que respaldan la selección de los límites de calificación



Con frecuencia, las distintas pruebas apuntan a los mismos resultados —como cuando los límites recomendados estadísticamente se encuentran en la zona de incertidumbre—, pero puede suceder que algunas se contradigan. En esos casos, es muy importante explicar la discrepancia al justificar la decisión final.

En esta última etapa del proceso de concesión de calificaciones finales, es posible mostrar qué efectos tendrían las distintas decisiones en los resultados totales del grupo de estudiantes. Teniendo en cuenta que lo importante es el resultado total y no los resultados de los componentes individuales, el equipo de concesión de calificaciones puede comparar la calificación media del grupo con las predicciones del profesorado, para lo cual consultará los resultados de años anteriores con el fin de comprender la fiabilidad de esa comparación. Otro enfoque posible consiste en excluir de los resultados totales a los nuevos colegios, para observar si obtienen resultados muy distintos a los de instituciones con más años de experiencia en el IB.

Finalmente, el equipo de concesión de calificaciones tiene que enviar sus recomendaciones para que se aprueben, acompañadas de las justificaciones necesarias. Entonces se realiza la totalización. La totalización es el proceso de combinar las puntuaciones y los límites de los distintos componentes para obtener una puntuación final o un límite de calificación global. Para ello, puede que sea necesario ajustar proporcionalmente las puntuaciones globales o los límites de los componentes. Para obtener más información, consulte la sección “Totalización” de esta publicación.

Límites de calificación fijos

En la mayor parte de la evaluación interna, y también en algunos componentes de evaluación externa, se pide al alumnado que realice tareas casi idénticas en cada convocatoria. Entre los ejemplos de esto se encuentran la preparación de una carpeta de Artes o el Proyecto Personal. En esos casos, es razonable dar por hecho que los límites de calificación no cambiarán, dado que el IB mantiene el mismo estándar de corrección año tras año (a través de la estandarización con el trabajo de años anteriores).

Durante el proceso de concesión de calificaciones finales, se pide al examinador/a principal y al examinador/a jefe que consideren si hay pruebas que indiquen que los límites de calificación existentes son

inadecuados. En la mayoría de los casos, se espera que la conclusión sea que se deben mantener los límites existentes.

A pesar de todo, hay que tener claro que los límites se verifican cada año, y que se pueden ajustar si hay pruebas que demuestren que ya no consiguen mantener la equiparabilidad de las calificaciones entre las convocatorias. Esto podría suceder en caso de que hubiera un cambio significativo en el comportamiento del alumnado (por ejemplo, debido a la influencia de avances tecnológicos en cómo se completa una tarea), cambios en los trabajos entregados que aumentasen la concordancia con el esquema de calificación pero sin aumentar la calidad general del trabajo del alumnado, o una revisión del enfoque del IB respecto a la corrección o a la moderación que modificase el estándar de corrección.

Para obtener más información sobre por qué cambian los límites de calificación, consulte el video [“El establecimiento de límites de calificación: por qué pueden cambiar y qué implica este cambio”](#).

Revisión y aprobación de las recomendaciones sobre los límites de calificación

Una vez que el equipo de concesión de calificaciones ha llegado a recomendaciones sobre dónde deben establecerse los límites de calificación, estas recomendaciones y sus justificaciones se presentan en el **informe de la asignatura para distribución en 24 horas**. Este informe también contiene datos subyacentes, como los cambios en el grupo de estudiantes que estudian el curso y las calificaciones finales otorgadas a partir de los límites propuestos, en comparación con años anteriores. Este informe lo revisan luego miembros directivos de la División de Evaluación del IB para determinar si los argumentos son lo suficientemente sólidos.

En caso de que el equipo directivo de la División de Evaluación del IB no esté conforme con las recomendaciones, discutirá sus inquietudes con la persona responsable de la asignatura y le pedirá que haga más hincapié en determinados aspectos de las pruebas o que ofrezca un mayor análisis para respaldar las recomendaciones.

El director/a en jefe de evaluación del IB tiene la última palabra sobre dónde se establecen los límites de calificación, basándose en las recomendaciones que recibe del equipo de concesión de calificaciones.

Otorgamiento de la titulación de un programa

Las titulaciones los programas del IB (el certificado del PAI, el diploma del IB y el certificado del POP) no se otorgan mediante un proceso de concesión de calificaciones. Su otorgamiento se determina durante la preparación para la publicación de los resultados, según el proceso que se describe en la sección [“Preparación para la publicación de los resultados”](#) de esta publicación.

Profesorado observador

El IB tiene el compromiso de aumentar la transparencia de sus procesos de evaluación y facilitar la comprensión general de cómo se conceden las calificaciones. Como parte de este compromiso, se invita al profesorado a asistir como observador a las reuniones de concesión de calificaciones (los detalles se concretarán en función de si son reuniones presenciales o virtuales). Se espera que el profesorado observador explique a sus compañeros/as lo que ha presenciado y ofrezca un informe para la comunidad de la Red de Educadores del IB (IBEN). Para obtener más información, comuníquese con [“El IB Responde”](#).

Principios del proceso de concesión de calificaciones finales

Los principios subyacentes del proceso de concesión de calificaciones finales del IB son los siguientes:

- Los límites de calificación 3-4, 6-7 y 2-3 se determinan (en ese orden) utilizando todas las pruebas disponibles (estimativas y estadísticas). En el caso de que no haya trabajos entregados para evaluación

para establecer estos límites de calificación, se utilizarán las pruebas disponibles para establecer el límite que resulte más adecuado.

- A continuación, los demás límites de calificación se determinan aritméticamente siguiendo el procedimiento adecuado.
- Las decisiones sobre dónde establecer los límites de calificación se toman a través de la triangulación de las pruebas obtenidas con las valoraciones de los examinadores/as, las pruebas estadísticas, y la información sobre el grupo de estudiantes. Todas estas pruebas se deben tener en cuenta por igual y se debe alcanzar un punto de equilibrio.
- Si el grupo de estudiantes de una evaluación es, en líneas generales, parecido al de años anteriores, los resultados también deberían ser similares a los de dichos años. Sin embargo:

Los grupos de estudiantes suelen variar de un año a otro, sobre todo en las asignaturas con poco alumnado.

Si los resultados varían respecto a los de años anteriores, se espera que haya pruebas claras que expliquen las razones.

Si las tareas de una evaluación son, en líneas generales, parecidas a las de años anteriores, los límites de calificación también deberían ser similares a los de dichos años.

Todos los límites de calificación pueden cambiar de un año a otro, incluso para las tareas de evaluación interna.

El IB procura que el nivel de dificultad de las evaluaciones sea coherente de un año a otro, pero dicho nivel puede variar en determinados exámenes.

La prioridad son los límites de calificación globales por asignatura, porque son los que afectan al alumnado y en los que se basan las partes interesadas para tomar decisiones.

Los límites de calificación por componente son una etapa esencial para obtener un resultado total sólido por asignatura, pero la acumulación de pequeños efectos a nivel de los componentes puede tener un efecto considerable en los resultados de la asignatura en su conjunto.

Controles de calidad

- Lo más importante que se obtiene de la evaluación es la calificación que recibe el alumnado; por ese motivo, los controles finales se centran en ella.
- El propósito de estos controles finales es buscar cualquier elemento inusual.

Después de establecer los límites de calificación, se realizan algunos controles de calidad adicionales para garantizar que las calificaciones que se conceden al alumnado sean justas y correctas.

Uno de estos controles es la revisión de la corrección de casos en riesgo. Esto identifica al alumnado con una calificación final inesperadamente baja. Parte de la revisión de la corrección de casos en riesgo se centra en el trabajo de examinadores/as cuyas correcciones se han identificado como merecedoras de una investigación más profunda. La revisión de la corrección la realizan examinadores/as que han demostrado que saben aplicar de forma coherente el estándar de corrección del examinador/a principal. Esta revisión de la corrección de casos en riesgo no se hace para identificar si se pueden conseguir algunos puntos más, sino para comprobar que la puntuación otorgada es adecuada; de este modo, se garantiza que el alumnado reciba calificaciones correctas.

El personal de evaluación del IB también realiza un control para verificar que las calificaciones de las asignaturas de los colegios parecen adecuadas.

El Comité de la Evaluación Final

- El Comité de la Evaluación Final proporciona un marco formal en el que el IB describe la convocatoria de exámenes para que la analicen varias partes interesadas invitadas que tengan relación con el PAI, el PD y el POP. El PEP no cuenta con un Comité de la Evaluación Final, ya que no hay resultados finales que verificar o certificar.
- El Comité de la Evaluación Final está formado por representantes de toda la organización, incluido personal sénior del IB y examinadores/as jefe. Además de dar su aprobación final para la publicación de resultados, también estudia casos de posibles infracciones de la política de integridad académica del IB y cualquier otra cuestión pertinente.

El Comité de la Evaluación Final representa la última etapa en la aprobación de los resultados del alumnado de los programas del IB que tienen evaluaciones creadas por la organización. Se trata del órgano que revisa las recomendaciones del director/a en jefe de educación y del director/a de evaluación del IB relativas a si se han respetado las normas y los estándares del IB durante la convocatoria de exámenes. Este comité concede los resultados en nombre del Consejo de Fundación del IB. Además, también es un foro para establecer políticas y precedentes relacionados con la concesión de las titulaciones del IB.

La composición y las atribuciones precisas del comité varían levemente en función del programa, pero en principio cuenta con la misma cantidad de miembros con derecho a voto, que se eligen de entre:

- Personal sénior del IB procedente de la Oficina de Educación y de la Oficina de Servicios a los Colegios
- Examinadores/as jefe de distintas áreas disciplinarias

Mientras que el personal sénior de evaluación es responsable de aprobar los límites de calificación, el Comité de la Evaluación Final actúa como órgano de supervisión. Este comité revisa los resultados a nivel global (por ejemplo, el porcentaje de estudiantes que completa cada programa) y cualquier cuestión que le presente el personal de la División de Evaluación del IB.

El comité también revisa las cuestiones de conducta impropia y mala administración, que debate en profundidad el Subcomité de Integridad Académica.

La función última del Comité de la Evaluación Final es reflexionar sobre el desempeño demostrado en la convocatoria de exámenes y aportar recomendaciones al IB para futuras convocatorias.

Conflicto de intereses

El Comité de la Evaluación Final es un órgano importante de toma de decisiones, por lo que es esencial que transmita una imagen de transparencia e independencia. Debido a esto, se aplica la siguiente política: si parece que un miembro puede tener un conflicto de interés en algún punto del orden del día, esa persona abandonará la sala mientras se trate ese tema. Esta medida de precaución se repite al principio de cada reunión.

Observadores

El IB fomenta que haya observadores/as en las reuniones del Comité de la Evaluación Final, con el fin de:

- Hacer que los procedimientos sean más transparentes
- Ofrecer la posibilidad de aportar sugerencias de cambios y de mejoras en los procesos
- Reconocer la colaboración entre las distintas partes que constituyen el IB

Para que el IB pueda cumplir estos objetivos, se pide a los observadores/as que envíen un informe sobre su participación al director/a en jefe de educación del IB dentro del plazo de dos semanas de realizarse la

reunión. En el informe se deben incluir observaciones generales sobre los procedimientos y los procesos actuales, y también sugerir cambios y mejoras cuando corresponda. No se deben hacer comentarios sobre casos individuales que haya abordado el comité.

Debido a lo delicado de las cuestiones que trata el comité, los observadores/as también están sujetos a las debidas restricciones relativas a la confidencialidad y a los conflictos de interés. Los observadores/as no tienen derecho a voto en las decisiones del comité.

Para obtener más información sobre cómo ser observador/a en una reunión del Comité de la Evaluación Final, escriba a support@ibo.org.

Preparación para la publicación de los resultados

- Una vez corregidas todas las evaluaciones y establecidos todos los límites de calificación, se deben completar varios procesos para que los resultados del alumnado estén listos para publicarse.
- En determinadas circunstancias, cuando no haya pruebas del trabajo entregado para evaluación, el IB puede calcular una puntuación mediante el procedimiento de ausencia de puntuación.
- Los resultados que cada alumno/a matriculado en un programa completo obtiene en las asignaturas se combinan para determinar si ha obtenido el certificado del PAI, el diploma del IB o el certificado del POP.
- La publicación de los resultados incluye un sistema eficaz de control de cambios. Cualquier alteración en las calificaciones se detecta, y se informa al respecto a las partes interesadas pertinentes, que principalmente son colegios y universidades.

Entre la etapa de establecimiento de los límites de calificación y la de publicación de los resultados para el alumnado hay varios procesos y procedimientos que se deben completar. Algunos de dichos procesos (como conservar las puntuaciones de quienes han cursado asignaturas en la categoría Anticipado) solo son necesarios para determinada parte del alumnado, mientras que otros pueden afectar a todo el alumnado.

Casos de ausencia de puntuación

- Cuando el IB no pueda acceder al trabajo de un alumno/a debido a motivos ajenos a su voluntad o la del colegio, se estimará una puntuación con el fin de minimizar los posibles efectos negativos para el alumno/a.
- Dicha estimación se debe basar en pruebas. Si los trabajos entregados para evaluación no son suficientes para realizar la estimación, entonces se debe encontrar otra solución para sustituir la puntuación ausente.
- El procedimiento de ausencia de puntuación se basa en los resultados medios del alumnado, por lo que esta estimación beneficiará aproximadamente al mismo número de estudiantes que a quienes perjudicará. El resultado más justo siempre será aquel que se obtenga al corregir el trabajo real entregado para evaluación.

En ocasiones, el IB no dispone del trabajo de un alumno/a y no lo puede corregir, por motivos que escapan del control del colegio y del alumno/a. Esto sucede, por ejemplo, cuando el servicio postal pierde los exámenes en papel con las respuestas del alumnado, o cuando un alumno/a contrae repentinamente una enfermedad el día del examen.

En tales situaciones, se puede emplear el procedimiento de ausencia de puntuación con el fin de estimar una puntuación para el alumno/a.

Este procedimiento está pensado para aquellas circunstancias en las que la ausencia del trabajo del alumno/a se debe a alguna acción del IB o de terceras partes (a excepción del colegio), a consecuencia de la cual no sería razonable exigir al alumno/a que completara la evaluación en otra ocasión.

El procedimiento de ausencia de puntuación siempre se debe basar en pruebas que demuestren los logros del alumno/a. Si el IB no dispone de demasiada información al respecto, es muy poco probable que se pueda estimar una puntuación justa.

El procedimiento de ausencia de puntuación se basa en los resultados medios que ha alcanzado todo el alumnado en el componente de evaluación afectado, que se comparan con los resultados obtenidos en los demás componentes. Al utilizar dicha media, se espera que la cantidad de estudiantes cuya puntuación

empeore debido a esta estimación sea aproximadamente la misma que de estudiantes cuya puntuación mejore. Por lo tanto, el resultado siempre será más justo si el IB puede corregir el trabajo real del alumno/a. Si en un colegio hay cinco estudiantes o más que cursen la asignatura en el mismo nivel, el IB utiliza los datos del colegio para determinar las puntuaciones ausentes. Si en un colegio hay menos de cinco estudiantes que cursen la asignatura, el IB utiliza datos mundiales para calcular las puntuaciones ausentes.

Figura 36

Cálculo de la puntuación ausente

Para determinar la puntuación ausente de un alumno/a:

$$\text{Cociente} = \frac{\text{Puntuación total ajustada del alumno en los otros componentes}}{\text{Puntuación total ajustada mundial (o del colegio) en los otros componentes}}$$

$$\text{Puntuación ausente del alumno} = \text{Puntuación ajustada mundial (o del colegio) en el componente que falta} \times \text{Cociente}$$

El cociente compara las puntuaciones del alumno/a con la media mundial o del colegio en los componentes que dicho alumno/a ha completado:

- Si el alumno/a ha obtenido mejores resultados que la media, esta cifra será mayor que 1.
- Si el alumno/a ha obtenido peores resultados que la media, esta cifra será menor que 1.

Al igual que sucede con todos los procedimientos de ausencia de puntuación, este cálculo es solo la mejor aproximación que se puede hacer a partir de las pruebas disponibles. Siempre es preferible poder corregir el trabajo real del alumnado.

Procedimiento de ausencia de calificación

En el PAI, si el modelo de evaluación de una asignatura concreta solo tiene un componente, entonces no se puede utilizar el procedimiento de ausencia de puntuación, ya que no habrá pruebas disponibles para calcular la puntuación. En esta situación, el IB utiliza un enfoque diferente: el procedimiento de ausencia de calificación del PAI.

Uno de los inconvenientes de querer reducir la carga de trabajo del alumnado es que, al utilizar solo un componente de evaluación, las pruebas disponibles en el PAI son escasas. Esto significa que el procedimiento de ausencia de calificación del IB es menos fiable que el procedimiento de ausencia de puntuación en los otros programas, por lo cual solo se debe utilizar en casos excepcionales.

Si un alumno/a ha completado la evaluación y la ha entregado tal como se requiere pero el IB no puede acceder a ella para su corrección debido a factores que escapan del control del alumno/a o del colegio, se aplica el procedimiento de ausencia de calificación.

Para que se pueda aplicar el procedimiento de ausencia de calificación, el alumno/a debe haber recibido una calificación final en al menos otros cuatro cursos del PAI. Es imposible calcular resultados para el alumnado que tenga menos calificaciones, ya que en esos casos no habrá suficientes pruebas como para realizar una estimación bien fundamentada.

Para determinar la calificación ausente, hay que calcular la calificación media en las otras asignaturas para las que exista una calificación obtenida a partir de trabajos entregados para evaluación en los últimos 18 meses.

Si la calificación media resultante tiene decimales superiores o iguales a 0,5, entonces se redondea al alza.

Si la calificación media resultante tiene decimales inferiores a 0,5, entonces se redondea a la baja.

Al igual que sucede con todos los procedimientos de ausencia de puntuación, este cálculo es solo la mejor aproximación que se puede hacer a partir de las pruebas disponibles. Siempre es preferible poder corregir el trabajo real del alumnado.

Resultados en los programas

- El IB otorga los diplomas y certificados de los programas siempre que se cumplan ciertos criterios precisos, y no basándose en valoraciones individuales.
- Cada programa se rige por criterios propios para la obtención del diploma o del certificado.

Después de conocer los resultados de un alumno/a en todas las asignaturas, se realiza un cálculo para determinar si puede recibir el certificado del PAI, el diploma del IB o el certificado del POP.

Para obtener información detallada sobre cómo se calculan los resultados en el PAI, el PD y el POP, así como los criterios para aprobar, consulte el [sitio web del IB](#) y los *Procedimientos de evaluación* del programa correspondiente.

Publicación de resultados

Desde la perspectiva del IB, la publicación de resultados exige establecer un protocolo muy estricto de control de cambios.

Durante el transcurso de la convocatoria de exámenes, la información se actualiza constantemente para poder saber con precisión qué está pasando. En cuanto se informa al alumnado y a los colegios sobre sus resultados, se debe indicar con claridad que no se realizará ningún cambio sin informar adecuadamente a todas las partes. En los casos del PD y del POP, esto es especialmente relevante para las universidades, que reciben los certificados de calificaciones finales del alumnado.

Después de la publicación de resultados se pueden autorizar cambios en estos debido a una consulta sobre los resultados, a la confirmación de resultados pendientes por parte de los colegios, o la resolución de problemas de integridad académica.

Consultas sobre los resultados y apelaciones sobre la evaluación

- El propósito de los servicios de consulta sobre los resultados es que los colegios puedan indicar al IB dónde creen que se ha producido un error durante el proceso de corrección.
- Durante las consultas sobre los resultados deben aplicarse los mismos estándares de corrección que durante la convocatoria de exámenes principal. Para asegurarse de que así sea, el IB utiliza el sistema de control previsto en el modelo de calidad.
- Las evaluaciones externas del IB están diseñadas para ser sumativas, y no formativas, así que se pide a los examinadores/as que solo escriban comentarios para explicar su corrección.
- El IB cuenta con un procedimiento formal de apelaciones que los colegios o el alumnado pueden emplear si consideran que los procesos no se han seguido correctamente.

Para conocer la descripción de los aspectos legales y de los procedimientos de los servicios de consulta sobre los resultados, consulte los *Procedimientos de evaluación* del programa correspondiente.

Categorías de consultas sobre los resultados

Tras la publicación de los resultados, el personal de coordinación puede realizar estos tipos de consultas sobre los resultados:

- **Consulta sobre los resultados de categoría 1:** Revisión de la corrección de todos los componentes de evaluación externa de un alumno/a en una asignatura
- **Informe de categoría 1:** Informe sobre la revisión de la corrección realizada a partir de una consulta sobre los resultados de categoría 1
- **Consulta sobre los resultados de categoría 2:** Copias de material de algún componente de evaluación externa.
- **Consulta sobre los resultados de categoría 3:** Revisión de la moderación de un componente de evaluación interna.

Los colegios deben abonar una tasa por cada una de las categorías anteriores, salvo cuando se modifique una calificación final como consecuencia de una consulta de categoría 1.

Propósito de las consultas sobre los resultados

El proceso de consulta sobre los resultados constituye la última defensa contra los errores que haya podido haber en el sistema de corrección. Se trata de una oportunidad para que los colegios indiquen al IB si creen que se ha producido algún error, y para que el IB realice las investigaciones necesarias y, si corresponde, corrija dichos errores.

Mantenimiento de estándares desde la corrección en condiciones reales hasta en las consultas sobre los resultados

El propósito de las consultas sobre los resultados es aportar al alumnado las puntuaciones que reflejen la calidad de su trabajo según los estándares del IB. Sin embargo, es muy importante que el personal examinador no se deje influir porque un alumno/a no esté satisfecho con su puntuación inicial, o porque dicha puntuación esté cerca de un límite de calificación.

En el proceso de consulta sobre los resultados, el IB emplea únicamente al personal examinador más coherente y con mayor nivel en la jerarquía. Por tanto, se considera que el resultado de esta revisión de la corrección es la puntuación correcta.

Para que el personal examinador mantenga el estándar de corrección adecuado, el IB incluye respuestas de control en el trabajo de consulta sobre los resultados. De este modo, el personal examinador recibe un aviso electrónico si se aleja del estándar establecido.

El personal de evaluación del IB revisa todas las propuestas de cambio de calificación tras la consulta sobre los resultados, con el fin de garantizar que la nueva calificación sea más fiable que la original. Si entre los dos examinadores/as existe una diferencia de puntuación muy notable que no se ha explicado, y se realiza un cambio de calificación, se puede pedir una tercera opinión sobre ese examen.

Consulta sobre los resultados de categoría 1: revisión de la corrección

Como se explica en la sección “[Corrección](#)”, dos examinadores/as pueden otorgar, para la misma pregunta, puntuaciones diferentes que se encuentren dentro de una diferencia aceptable (margen de tolerancia). Por lo tanto, si al concluir el proceso de consulta sobre los resultados se determina que se debe realizar tan solo un leve cambio en la puntuación, es probable que no haya habido ningún error de corrección. Esto es lo que suele suceder en la mayoría de las solicitudes de consulta sobre los resultados.

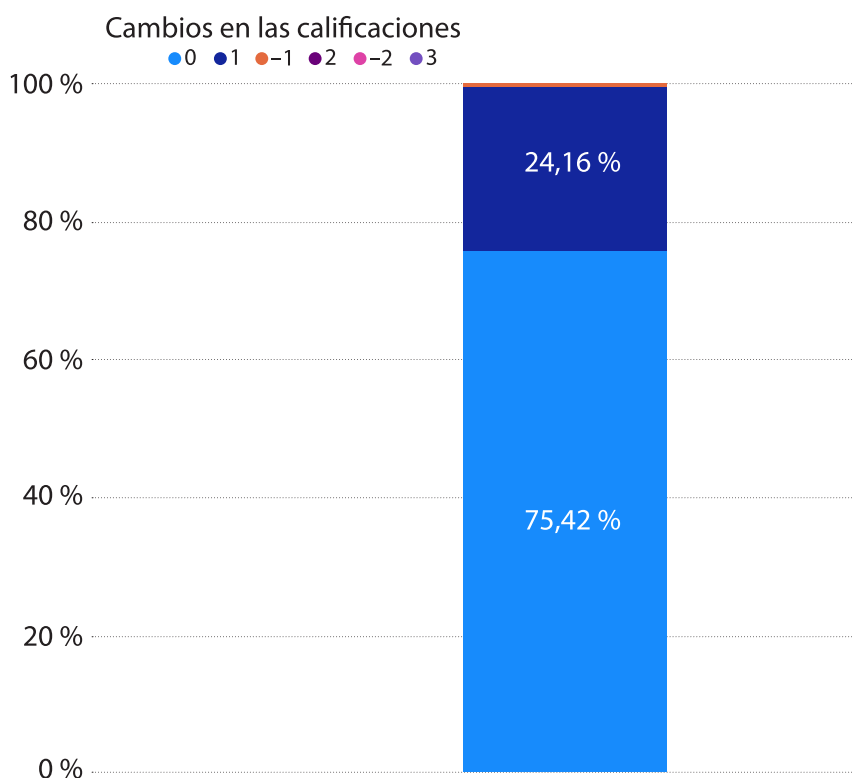
A fin de minimizar los efectos de estas diferencias menores —aunque aceptables— entre las puntuaciones del personal examinador, durante una consulta sobre los resultados de categoría 1 se revisa la corrección de todas las evaluaciones externas del alumno/a. De este modo se mitiga el impacto de los pequeños cambios, ya que, siempre que no sean resultado de problemas sistemáticos, se deberían compensar entre sí a lo largo de distintos exámenes.

El IB considera que la puntuación que se decide tras una consulta sobre los resultados es la representación más fiel del trabajo del alumno/a, por lo que esa será la puntuación final que se otorgará. Esto significa que, tras una consulta sobre los resultados de categoría 1, las puntuaciones del alumno/a pueden aumentar o disminuir. Por ello, es obligatorio que los colegios pidan al alumno/a permiso para que su trabajo pase por este procedimiento.

Teniendo en cuenta la naturaleza del proceso para establecer los límites de calificación, siempre habrá estudiantes que se queden tan solo 1 punto por debajo del límite, o por encima. En tales casos, la consulta sobre los resultados podría provocar que hubiera un cambio en la calificación, aunque esto no implica necesariamente que haya habido algún error en la corrección. En este tipo de situación, el IB sostendría que el alumno/a se encuentra en el límite entre dos calificaciones, y que ambas reflejan de manera justa su desempeño.

Figura 37

Media de cambios en las calificaciones debido a consultas sobre los resultados de categoría 1 a lo largo de cinco convocatorias de exámenes del PD



Consulta sobre los resultados de categoría 2: devolución de materiales

Las evaluaciones que pueden optar a una consulta sobre los resultados de categoría 2 son las evaluaciones sumativas. Un examen devuelto puede contener comentarios útiles del examinador/a a quien se le había asignado. Sin embargo, no se puede garantizar que sea así, ya que el personal examinador no está obligado a escribir comentarios cuando corrige los trabajos.

Se espera que el personal examinador corrija de manera adecuada, para garantizar que el alumnado obtenga resultados justos que reflejen el trabajo que haya entregado. Solamente se le pide que incluya comentarios para explicar su corrección, y no aportar comentarios al alumnado para indicarle cómo podría mejorar.

Para obtener más información sobre qué hacer si se considera que un trabajo no se ha corregido de manera justa, consulte el video ["Resultados problemáticos: qué hacer si cree que su trabajo no se ha calificado correctamente"](#).

Consulta sobre los resultados de categoría 3: revisión de la moderación

En lo que respecta a las consultas sobre los resultados de categoría 3 (revisión de la moderación), cualquier cambio en la puntuación afectará a todo el grupo de estudiantes. Si en estos casos se aplicara también el principio de contar con el permiso del alumnado, esto implicaría que si un solo alumno/a no diera su aprobación o simplemente no respondiera a esta solicitud, podría impedir la revisión de la moderación de toda su clase. Sin duda, esto no sería razonable. Por tanto, para evitar que disminuyan las calificaciones del alumnado sin que hayan aceptado formalmente este riesgo, las consultas sobre los resultados de categoría 3 **solo pueden aumentar** las calificaciones, y nunca disminuirlas.

La revisión de la moderación solo se realizará en los casos en los que el promedio de las puntuaciones de evaluación interna moderadas difiera del promedio de las originales otorgadas por el profesor/a en **al menos un 15 %** de la puntuación máxima del componente.

Este umbral se ha establecido porque si el IB no impone ninguna restricción, algunos colegios pedirían sistemáticamente revisiones de la moderación de todas sus asignaturas con la esperanza de que suban algunas calificaciones. La cantidad limitada de personal moderador supervisor del IB no permitiría hacer frente a este nivel de demanda. Al fijar este parámetro para aceptar la revisión de la moderación se garantiza que solo los colegios con una clara inquietud puedan acceder al servicio. De este modo, el IB puede gestionar el volumen de trabajo de su poco personal moderador supervisor y ofrecer, al mismo tiempo, revisiones de la moderación a los colegios cuya moderación ha afectado considerablemente a las puntuaciones del alumnado.

Figura 38

Grupo de estudiantes que cumple los requisitos para una consulta sobre los resultados de categoría 3

Número del alumno/a	¿Es una muestra?	Puntuaciones sin moderar	Puntuaciones moderadas	Ajuste de moderación
2	Sí	19	14	-5
10		15	12	-3
11	Sí	16	13	-3
12		20	15	-5
13		18	14	-4
16	Sí	12	11	-1
36	Sí	17	13	-4
45	Sí	10	10	0
Cambio promedio como puntuación				-3,125
Cambio promedio como porcentaje de la puntuación máxima				-15,63 %

En el ejemplo que se muestra en la figura 38, la puntuación máxima obtenida en el componente es 20.

La media del ajuste de moderación es -3,125 puntos, lo que equivale al 15,63 % de la puntuación máxima.

Esto significa que la muestra cumple los requisitos para una revisión de la moderación de categoría 3.

Apelaciones sobre la evaluación

Una apelación es una oportunidad para solicitar al IB la revisión de cómo se tomó una decisión durante el proceso de evaluación, por ejemplo, sobre el procesamiento de la calificación final del IB otorgada a un alumno/a o sobre la gestión de un caso de conducta impropio de un alumno/a.

A través de una apelación no se podrá solicitar una excepción a un proceso del IB, pero sí cuestionar si la interpretación de las normas del IB es razonable o no. Las políticas publicadas por el IB no podrán modificarse a raíz de una apelación. Asimismo, cuando una apelación dé lugar a una aclaración con respecto a las normas del IB, esta deberá aplicarse a la totalidad del alumnado de la convocatoria en cuestión. Una apelación supone el último paso en los casos de dudas sobre el proceso de evaluación del IB.

Para obtener más información, consulte la publicación *Procedimientos de evaluación* del PAI, del PD o del POP, disponibles en el Centro de Recursos para los Programas.

Alumnado de la categoría Repetidor

Si un alumno/a no está satisfecho con la calificación final obtenida, puede repetir los exámenes de la asignatura en cuestión en la convocatoria que se celebra seis meses después o, en principio, en cualquier convocatoria posterior. Sin embargo, si se han introducido nuevos requisitos curriculares o de evaluación significativos cuando se vaya a repetir la asignatura, el alumno/a deberá cumplir con tales requisitos. Un alumno/a que desee repetir una o varias asignaturas no necesariamente tiene que matricularse en el mismo colegio en el que cursó dichas asignaturas con anterioridad. Cuando un colegio admite a un alumno/a que desea repetir una o varias asignaturas, debe asumir todas las responsabilidades académicas y administrativas relacionadas con dicho alumno/a.

Los alumnos/as que se presenten de nuevo a una asignatura pueden conservar sus resultados de un componente que no sea un examen, suponiendo que el programa de estudios y los requisitos de evaluación no hayan cambiado de manera significativa. No se pueden conservar las puntuaciones de los exámenes escritos, por lo cual el alumno/a tendría que volver a realizar todos los exámenes correspondientes a la asignatura en cuestión.

Si un alumno/a desea volver a presentar trabajos para cualquier componente que no sea un examen, deberá asistir a clases en el colegio donde está matriculado para la convocatoria con la categoría Repetidor. La razón de ello es que el profesor/a de la asignatura debe proporcionar orientación académica, corregir el trabajo de evaluación interna y confirmar la autoría original de todo el trabajo. No es suficiente que el alumno/a realice pequeñas modificaciones al trabajo que haya entregado anteriormente. En principio, debe presentarse un trabajo totalmente distinto para la evaluación. No obstante, el IB es consciente de que este requisito puede no ser factible cuando se trata de tareas de cierta envergadura dadas las limitaciones de tiempo. En tales casos se aceptarán modificaciones sustanciales del trabajo original. Si se presenta un trabajo modificado, se evaluará como si fuera nuevo. El alumno/a deberá aceptar que el trabajo puede recibir una puntuación más baja. En el caso de los trabajos modificados, el alumno/a debe haber dedicado tiempo adicional equivalente, como mínimo, al 20 % del total de horas dedicadas al trabajo original. Además, el trabajo debe cumplir al menos uno de los siguientes criterios:

- Se ha cambiado la pregunta de investigación, el enfoque o la tesis.
- Se ha añadido un tema, aspecto de las pruebas o punto de vista nuevo que influye en las conclusiones.
- Se han llevado a cabo un análisis o cálculos adicionales significativos.
- Se ha explorado un aspecto nuevo (adicional) de la obra de un autor/a o artista.

En los casos en que el alumno/a tenga trabajo completado que no se haya entregado para que forme parte de su evaluación (por ejemplo, carpetas de Artes del PD), se permite computar el tiempo ya dedicado a crear este trabajo, que no se había visto anteriormente, para alcanzar el objetivo del 20 %.

Se elige este enfoque teniendo en cuenta dos aspectos. El primer aspecto es la expectativa de que el alumnado de la categoría Repetidor no se beneficie de otra ronda de comentarios sobre su trabajo, lo cual pondría en desventaja a los alumnos/as que presentan trabajos nuevos. El segundo aspecto es que no sería equitativo imponer una carga excesiva de trabajo al alumnado o al colegio para obtener, por ejemplo, datos o investigaciones totalmente nuevos que, a efectos prácticos, impedirían realizar de nuevo la evaluación.

Para obtener más información sobre el enfoque del IB acerca de la equiparabilidad, consulte la sección [“Equiparabilidad”](#) de esta publicación.

El IB es consciente de que, con el tiempo, será necesario modificar el contenido de un curso. Antes de introducir un nuevo curso, el IB comunica cualquier cambio a los colegios de diversas maneras, entre ellas, a través de la publicación sobre novedades del programa pertinente y la elaboración de una nueva guía de la asignatura. Después de la última convocatoria de exámenes del curso anterior, el IB no puede ofrecer al alumnado de la categoría Repetidor la oportunidad de realizar exámenes relacionados con dicho curso anterior. Esto se debe, en parte, a que sería inviable tanto para los colegios como para el IB que se evaluase al alumnado en dos cursos diferentes dentro de la misma asignatura en la misma convocatoria de exámenes. Sin embargo, es posible que el alumnado de la categoría Repetidor pueda conservar sus

puntuaciones de evaluación interna o las de los trabajos de clase evaluados externamente. Si el nuevo curso es sustancialmente diferente al anterior, es poco probable que esta opción esté disponible.

Para obtener más información sobre el alumnado de categoría Repetidor, consulte la publicación *Procedimientos de evaluación* del programa correspondiente.

Comentarios para los colegios

- El propósito de la evaluación sumativa del IB es medir el desempeño del alumnado, y todos los procesos pertinentes se diseñan de tal modo que permitan obtener los resultados más válidos que sea posible.
- El IB prioriza la transparencia en la comunicación de los resultados de la evaluación a los colegios y al alumnado, con el fin de que estudiantes y docentes vean que las puntuaciones y las calificaciones se otorgan de manera coherente y sin sesgos.
- La transparencia es de especial importancia en la evaluación interna, ya que el resultado se basa en una moderación de las puntuaciones del profesorado.
- La finalidad de la evaluación sumativa del IB, incluido el proceso de consulta sobre los resultados, no es ofrecer orientación a los colegios sobre cómo mejorar los resultados del alumnado, a menos que esto sea una consecuencia indirecta de los datos necesarios para una evaluación válida.
- Aparte del proceso de evaluación, el IB proporciona a los colegios otros servicios de apoyo pedagógico y de desarrollo profesional.

El propósito de la evaluación sumativa del IB es sopesar el desempeño del alumnado una vez concluido el programa de estudios y proporcionar una calificación que sirva como medición fiable de dicho desempeño. Esto es diferente de la evaluación formativa, cuyo propósito es proporcionar comentarios para ayudar al alumnado en su aprendizaje. Si bien aportar comentarios significativos forma parte esencial de la evaluación formativa y de la enseñanza eficaz, no es apropiado utilizar los resultados de las evaluaciones sumativas con fines formativos. Un motivo de esto es que un examinador/a del IB ofrecerá su valoración a partir del trabajo que tiene a su disposición (es decir, el examen con las respuestas del alumno/a), mientras que, en el aula, el profesorado del IB puede utilizar una amplia variedad de información y su conocimiento personal para decidir qué comentarios aportar a un alumno/a. Por consiguiente, el personal examinador no proporciona comentarios escritos para el alumnado ni para el profesorado al corregir.

Los examinadores/as deben indicar claramente en qué partes otorgan puntos y, en caso de ambigüedad, explicar su decisión mediante comentarios adecuados. Esto permite al IB comprobar los estándares de corrección y hace que el proceso de concesión de puntuaciones sea más transparente para los colegios.

Informes generales de las asignaturas

Después de cada convocatoria de exámenes, el IB publica un informe general de cada asignatura, que normalmente elaboran los examinadores/as jefe. El informe general de la asignatura proporciona información sobre cada componente de evaluación y sobre el desempeño del alumnado en los exámenes, y se indican las preguntas o los temas que se han abordado especialmente bien o especialmente mal. El informe contiene recomendaciones generales de los equipos examinadores sobre cómo el profesorado puede preparar a sus estudiantes para abordar preguntas o temas similares en futuras convocatorias. En los informes generales de las asignaturas también se proporcionan los límites de calificación para la asignatura, así como los límites de cada componentes individual.

Lógicamente, cada docente tendrá que adaptar la información que se da en el informe general de la asignatura al contexto de su propia clase. Por ejemplo, aunque una pregunta concreta haya obtenido respuestas mediocres en general, es posible que todo el alumnado de una clase haya obtenido puntuaciones altas en dicha pregunta.

Informes sobre la evaluación interna

Las puntuaciones que otorga el profesorado en la evaluación interna están sujetas a un proceso de moderación. Es posible que el profesorado interprete los criterios de evaluación de la evaluación interna de manera diferente a un examinador/a del IB. Por ejemplo, puede que corrijan el trabajo de forma más estricta y que otorguen menos puntos. El propósito de la moderación de la evaluación interna es ajustar el estándar de corrección del profesorado para que concuerde con el del examinador/a principal o el estándar global. Para ayudar al personal docente a comprender en qué varía su corrección con respecto al estándar global, el IB proporciona comentarios sobre la corrección de las tareas de evaluación interna. Esto sucede cuando las puntuaciones que concede el profesorado quedan fuera de los márgenes de tolerancia, y permite al profesorado entender por qué se ha aplicado un factor de moderación.

Sin embargo, debe quedar claro que los comentarios que se aportan a los colegios no tienen por objeto explicar cómo el alumnado de la muestra podría haber conseguido mejores resultados. Más bien, ofrece información sobre la calidad de la corrección de cada criterio por parte del profesorado, así como sobre su aplicación general de los criterios de evaluación y la idoneidad de los trabajos entregados.

Nota sobre los libros de texto, los talleres y los exámenes

El currículo de cada uno de los cursos del IB se detalla en la guía de la asignatura correspondiente y sirve de base para diseñar las evaluaciones. Los libros de texto, incluidos los que cuentan con la aprobación del IB, pretenden servir de ayuda para que estudiantes y docentes completen el curso tal como se establece en la guía de la asignatura, y no definir el contenido del currículo.

Por ese motivo, las preguntas de examen pueden tratar sobre temas, o partes de temas, que figuren en la guía de la asignatura pero no en un libro de texto concreto. En ese sentido, se insta a todo el profesorado a remitirse exclusivamente a las guías de las asignaturas para toda cuestión referida al contenido del currículo. Como parte del proceso de redacción de los exámenes, el IB comprueba los libros de texto que se utilizan más frecuentemente, con el fin de asegurarse de que las preguntas que figuren en dichos libros no aparezcan en los exámenes.

Asimismo, los comentarios y los documentos que se proporcionan en los talleres del IB no sustituyen al contenido de las guías de las asignaturas, aunque sí pueden ayudar al profesorado a interpretar mejor dicho contenido. Si se realizan modificaciones a las guías, el IB publicará formalmente estos cambios y los describirá claramente como tales.

Definición de los procesos específicos de los programas

- Puesto que todos los programas del Bachillerato Internacional (IB) se basan en la misma filosofía educativa, todas las evaluaciones del IB deben basarse en los mismos principios y prácticas generales para respetar dicha filosofía. Los propósitos de la evaluación cambiarán de manera natural a medida que el alumnado avance por los distintos programas en las diferentes etapas de la maduración educativa.
- El propósito general de contar con evaluaciones sumativas externas en los programas del IB es proporcionar al alumnado un resumen válido de sus logros académicos que les ayude a avanzar en su educación o en el mundo laboral.
- Si bien la evaluación es un elemento importante del Programa de la Escuela Primaria (PEP), no resulta adecuado que cuente con una evaluación sumativa externa.
- Cada programa tiene algunas características distintivas que requieren procesos específicos dentro del marco de las prácticas de evaluación.

Elementos comunes a todos los programas

- Los programas del IB ofrecen currículos o marcos curriculares amplios, equilibrados, conceptuales e interconectados.
- Al diseñar las evaluaciones del IB, todas deben incluir estos aspectos propios de una educación del IB, aunque no se aborden explícitamente, para que haya un efecto de repercusión positivo en el aprendizaje y la enseñanza.

Figura 39

Continuo de programas del IB



En la publicación *¿Qué es la educación del IB?* (2019) se hace hincapié en que los programas del IB promueven el aprendizaje conceptual, que se centra en ideas organizadoras amplias e importantes que son pertinentes dentro de las distintas áreas disciplinarias y entre ellas, que ayudan a integrar el aprendizaje, y que contribuyen a la coherencia del currículo. Los programas hacen hincapié en la importancia de establecer conexiones, explorar las relaciones entre las disciplinas académicas y aprender acerca del mundo en formas que van más allá del ámbito de cada disciplina. Los programas ofrecen al alumnado acceso a una variedad de estudios académicos y experiencias de aprendizaje que son amplios, equilibrados y conceptuales y están interconectados.

- **Amplios y equilibrados:** La educación del IB representa un enfoque equilibrado que ofrece al alumnado acceso a una amplia gama de contenidos que abarcan las disciplinas académicas.
- **Conceptuales:** El aprendizaje conceptual se centra en ideas organizadoras amplias e importantes, pertinentes dentro de las distintas áreas disciplinarias y entre ellas, y con un alcance que trasciende fronteras nacionales y culturales. Los conceptos ayudan a integrar el aprendizaje, contribuyen a la coherencia del currículo, profundizan la comprensión disciplinaria, desarrollan la capacidad de abordar ideas complejas y permiten la transferencia del aprendizaje a nuevos contextos.
- **Interconectados:** El marco curricular del IB valora la simultaneidad del aprendizaje. El alumnado entra en contacto con muchas disciplinas simultáneamente a lo largo de sus programas de estudio; aprende a establecer conexiones y enriquece su comprensión de la interrelación entre el conocimiento y la experiencia en muchas áreas. Los objetivos generales de los cursos y los requisitos de los programas ofrecen oportunidades auténticas para aprender acerca del mundo en formas que van más allá del ámbito de cada disciplina.

Cuando diseñen evaluaciones individuales, el profesorado, los colegios y los autores/as del IB tienen que reflexionar sobre estos objetivos subyacentes para no crear tareas que obstaculicen un buen aprendizaje y una buena enseñanza. El propósito siempre debe ser generar un efecto de repercusión positivo.

La dimensión internacional

Los principios del IB apuntan a ayudar al alumnado a convertirse en “personas solidarias [...] capaces de contribuir a crear un mundo mejor y más pacífico, en el marco del entendimiento mutuo y el respeto intercultural” y a alentarle a “entender que otras personas, con sus diferencias, también pueden estar en lo cierto” (Declaración de principios del IB, 2024). El alumnado que estudia los programas del IB se encuentra en numerosos países y regiones o territorios, y tiene una amplia variedad de nacionalidades. Por lo tanto, en el aprendizaje y la enseñanza del IB hay un contexto internacional y un propósito de desarrollar una mentalidad internacional, y ambos elementos deben reflejarse en la evaluación. Para obtener más información sobre el enfoque del IB acerca de la mentalidad internacional, consulte la sección [“Mentalidad internacional y entendimiento intercultural”](#) de esta publicación.

El perfil de la comunidad de aprendizaje del IB

Actualmente, la educación trata mucho más sobre formas de pensar que implican enfoques creativos y críticos para la resolución de problemas y la toma de decisiones. Asimismo, aborda métodos de trabajo, en particular de comunicación y colaboración, así como las herramientas necesarias, tales como la capacidad de reconocer y explotar el potencial de las nuevas tecnologías, y también de evitar sus riesgos. Por último, la educación abarca la capacidad de vivir como ciudadano activo y comprometido en un mundo polifacético. Estos ciudadanos influyen en lo que quieren aprender y cómo quieren hacerlo, y eso da forma al papel de los educadores.

(Schleicher, 2016)

Hay muchas maneras distintas de categorizar estas habilidades y competencias, por ejemplo, el marco de competencias para el siglo XXI de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), el que propone RAND Education, y el del National Research Council (NRC). En el IB, estas competencias se describen mediante el perfil de la comunidad de aprendizaje.

Figura 40

El perfil de la comunidad de aprendizaje del IB



No todos los aspectos del perfil de la comunidad de aprendizaje se pueden medir adecuadamente con la evaluación sumativa. El aprendizaje, la enseñanza y la evaluación de buena calidad reconocen la importancia de estas características y —aunque no estén diseñados para medirlas— pueden ofrecer al alumnado la oportunidad de desarrollar estas competencias. Algunos ejemplos serían fomentar enfoques éticos (íntegros) para las encuestas y los experimentos, respaldar una revisión adecuada entre compañeros/as, e introducir contextos inesperados para el alumnado.

Lecturas adicionales

Para obtener más información sobre los valores en los que se basa el sistema educativo del IB, puede consultar los siguientes recursos:

- *¿Qué es la educación del IB?*
- El perfil de la comunidad de aprendizaje del IB
- *De los principios a la práctica* del Programa de los Años Intermedios (PAI), del Programa del Diploma (PD) y del Programa de Orientación Profesional (POP)

Necesidades y soluciones específicas de los programas

En secciones anteriores de esta publicación se han explicado los procesos que conforman el ciclo de evaluación. Esas prácticas generales sirven para todos los programas del IB. Sin embargo, cada programa tiene un propósito distinto, así que habrá diferencias en algunos aspectos de la evaluación de cada uno de ellos y en las reglas que determinan si se concede la titulación de un programa.

Estos enfoques de la evaluación, específicos para cada programa, son importantes. En esta sección se explica en qué consisten y cómo se gestionan.

Programa del Diploma

Las características distintivas de la evaluación sumativa externa del Programa del Diploma (PD) son:

- El alumnado debe cursar una cantidad obligatoria de asignaturas para obtener el diploma del IB.
- El logro total alcanzado en el diploma del IB se describe mediante una puntuación cuyo máximo es de 45 puntos.
- Los componentes troncales aportan hasta 3 puntos al resultado total en el diploma del IB, a través de una matriz de puntos.
- Casi todas las asignaturas tienen varios componentes de evaluación, y combinan la evaluación externa y la evaluación interna.
- Casi todas las asignaturas están disponibles en Nivel Medio (NM) y Nivel Superior (NS), y contribuyen a partes iguales al resultado global en el diploma del IB.

Objetivos generales del Programa del Diploma

Los resultados de la evaluación solo son válidos si están claramente relacionados con los objetivos generales del curso y del programa.

El principal objetivo general del PD es ofrecer a estudiantes de entre 16 y 19 años de edad una experiencia educativa estimulante con un enfoque internacional, amplio y equilibrado mediante el estudio simultáneo de seis asignaturas y una serie de componentes troncales durante dos años. Otro objetivo general del PD es que el alumnado desarrolle las habilidades académicas básicas que requerirá en sus estudios universitarios, sus estudios posteriores y la profesión que elija. Además, el programa apoya el desarrollo de las habilidades prácticas y los valores para una vida plena y fructífera. Dicho de otro modo, la validez de los resultados de la evaluación en el PD, sus cursos y componentes troncales se basa en la atención a la totalidad de la experiencia educativa de cada estudiante.

Usos válidos de los resultados de las evaluaciones del PD

Cuando el IB desarrolla los modelos de evaluación y los currículos, procura que las calificaciones finales obtenidas en los cursos del PD y la puntuación total obtenida en el diploma del IB se puedan utilizar para determinar los siguientes aspectos:

- La selección para el ingreso a la universidad o para oportunidades laborales
- Si el alumno/a ya ha cumplido algunos de los requisitos de algún programa universitario, con el posible reconocimiento de ello mediante el otorgamiento de créditos para obtener una titulación o la validación de algunos cursos
- Prueba de que un alumno/a es capaz de cursar estudios posteriores una lengua determinada

Para obtener más información sobre la importancia de estos usos válidos, consulte la sección “Validez” de esta publicación.

La estructura del PD y sus cursos

Para obtener el diploma del IB, el alumnado debe cursar seis asignaturas junto con los componentes troncales: Teoría del Conocimiento (TdC), la Monografía, y Creatividad, Actividad y Servicio (CAS).

Se deben cursar algunas asignaturas del Nivel Medio (NM) y otras del Nivel Superior (NS). Cada estudiante debe cursar un mínimo de tres asignaturas (y un máximo de cuatro) del NS, y el resto del NM. Los cursos del NM y el NS se diferencian en cuanto a su alcance, pero el desempeño en estos está en consonancia

gracias a que comparten los mismos descriptores de calificaciones finales y objetivos de evaluación de los grupos de asignaturas. El alumnado debe demostrar un grado más amplio de conocimientos, comprensión y habilidades en el NS.

También existen cursos interdisciplinarios que cubren más de un grupo de asignaturas. Por ejemplo, Sistemas Ambientales y Sociedades, que cumple el requisito de estudiar un curso del grupo de Individuos y Sociedades y el de Ciencias.

Figura 41

Modelo del Programa del Diploma



Modo de calcular el resultado del PD

La validez del PD se refleja en la forma en que se mide el desempeño de cada estudiante. Para calcular la puntuación total obtenida en el diploma del IB, se suman las calificaciones (del 1 al 7) obtenidas en cada una de las seis asignaturas, y después se añaden entre 0 y 3 puntos por el desempeño en los componentes troncales (TdC, la Monografía, y CAS). Las asignaturas de NM y NS tienen el mismo peso a la hora de determinar la puntuación final del alumno/a. La puntuación más alta posible es 45 puntos.

Matriz de puntos para los componentes troncales

A diferencia de las asignaturas, TdC y la Monografía reciben calificaciones alfabéticas, que van de la A a la E. La naturaleza del tercer componente troncal, CAS, implica que no se le asigne una calificación.

El desempeño en los componentes troncales contribuye entre 0 y 3 puntos a la puntuación total obtenida en el diploma del IB. Obtener una calificación de E en TdC o en la Monografía, o no completar los requisitos de CAS, implica que el alumno/a no podrá recibir el diploma del IB. La puntuación que se concede en los componentes troncales se calcula tal como se muestra en la figura 42.

Figura 42

Concesión de puntos para TdC y la Monografía

		Teoría del Conocimiento (TdC)					
		Calificación concedida	A	B	C	D	E o N
Monografía	A	3	3	2	2	No puede recibir el diploma del IB	
	B	3	2	2	1	No puede recibir el diploma del IB	
	C	2	2	1	0		
	D	2	1	0	0		
	E o N	No puede recibir el diploma del IB					

Requisitos para obtener el diploma del IB

Un alumno/a puede recibir el diploma del IB si se cumplen las siguientes condiciones:

- Haber cumplido los requisitos de CAS
- Haber obtenido una puntuación total de 24 como mínimo
- No haber obtenido una N (no se ha concedido ninguna calificación) en TdC, la Monografía o alguna de las asignaturas que cuentan para la obtención del diploma
- Haber obtenido como mínimo una calificación D en TdC y la Monografía
- No haber obtenido una calificación de 1 en ninguna asignatura o nivel
- No haber obtenido una calificación de 2 tres o más veces, ya sea en el NM o el NS
- No haber obtenido una calificación de 3 o inferior cuatro o más veces, ya sea en el NM o el NS
- Haber obtenido 12 puntos o más en las asignaturas de NS (para el alumnado matriculado en cuatro asignaturas de NS, contarán las tres calificaciones más altas)
- Haber obtenido 9 puntos o más en las asignaturas de NM (el alumnado matriculado en dos asignaturas de NM deberá obtener al menos 5 puntos en este nivel)

Diploma bilingüe

En lugar del diploma del IB habitual, un alumno/a puede obtener un diploma bilingüe si cumple **al menos uno** de los siguientes criterios:

- Haber completado dos cursos de lengua de Estudios de Lengua y Literatura, con una calificación mínima de 3 en cada uno de ellos.
- Haber completado una asignatura de Individuos y Sociedades o de Ciencias en una lengua distinta a la cursada en Estudios de Lengua y Literatura. El alumno/a debe haber obtenido como mínimo una calificación de 3 tanto en la lengua de Estudios de Lengua y Literatura como en la asignatura de Individuos y Sociedades o de Ciencias.

Las asignaturas piloto y las asignaturas interdisciplinarias contarán para la obtención de un diploma bilingüe, siempre que se cumplan las condiciones anteriores.

No contarán para la obtención de un diploma bilingüe:

- Una monografía
- Una asignatura del programa de estudios del colegio
- Una asignatura adicional a las seis asignaturas conducentes al diploma del IB

Gestión de los estándares en grupos de estudiantes de distintos tamaños

En las asignaturas del PD, los grupos de estudiantes pueden tener un tamaño muy dispar, desde solo 1 estudiante hasta más de 40.000. Aunque todas las asignaturas son igual de importantes, por razones prácticas, se utilizan procesos ligeramente distintos en función del número de estudiantes

Mantenimiento de la calidad del trabajo del personal examinador

En las asignaturas en las que hay muy pocos examinadores/as, puede que no resulte adecuado utilizar el [modelo de calidad](#) (con respuestas de práctica, de aptitud y de control) que se describe en la sección “Estandarización” de esta publicación.

Si todos los examinadores/as participan en la discusión que se realiza en el momento de la estandarización, no hará falta que haya respuestas de práctica ni de aptitud, puesto que todos los examinadores/as ya habrán participado en el proceso de establecer del estándar. No obstante, se seguirán utilizando respuestas de control para garantizar que los examinadores/as continúan corrigiendo según el estándar previsto. Sin embargo, normalmente se prepara un modelo de calidad completo, aunque no se necesite.

Lo importante es mantener la calidad de la corrección, aunque una asignatura tenga muy poco alumnado.

Calificaciones: proceso completo de concesión de calificaciones (virtual o presencial)

En el caso de las asignaturas en las que hay varios examinadores/as por debajo del examinador/a jefe o del examinador/a principal, se realiza una reunión formal para aplicar las prácticas que se describen en la sección “[Concesión de calificaciones finales y totalización](#)” de esta publicación. En esta reunión habrá una persona responsable de la asignatura del IB para ayudar a los examinadores/as y realizar un control de calidad del proceso. Estas reuniones pueden ser presenciales o virtuales.

Calificaciones: proceso guiado de concesión de calificaciones en asignaturas con poco alumnado (virtual)

En los casos en los que hay muy pocos examinadores/as en cada componente, los examinadores/as supervisores se reúnen virtualmente para discutir. La persona responsable de la asignatura del IB les dará toda la gama de pruebas necesarias para decidir adecuadamente las calificaciones que se deben conceder, y también se encargará de supervisar el proceso para garantizar la calidad.

Calificaciones: proceso clásico de concesión de calificaciones en asignaturas con poco alumnado

Cuando el número de estudiantes es demasiado bajo como para que las pruebas estadísticas acumuladas con los años sean significativas, se pide al equipo examinador que revise todos los trabajos entregados para evaluación en vez de revisar solo una muestra para recomendar los límites de calificación estimativos.

Los/as responsables de las asignaturas del IB prestan apoyo en estas reuniones según se les solicite, y su función no es observar las reuniones, sino realizar controles de calidad aleatorios.

Para facilitar el trabajo de los examinadores/as de todas las asignaturas con poco alumnado (ya sea en el proceso guiado o en el proceso clásico de concesión de calificaciones), se recomienda realizar discusiones entre los examinadores/as de las distintas asignaturas de un mismo grupo. Esto contribuye a garantizar que haya una comprensión común de lo que significan las calificaciones en todo el grupo de asignaturas.

Aprobación final

Independientemente de la cantidad de estudiantes que hayan realizado los exámenes, la última etapa del proceso de concesión de calificaciones es que el examinador/a jefe o el examinador/a principal haga recomendaciones al personal de evaluación sénior del IB, y que este confirme que las pruebas que respaldan esas recomendaciones le convencen.

Independientemente del número de estudiantes en un grupo, todas las asignaturas se someten al mismo escrutinio.

Lecturas adicionales

Para obtener más información sobre el PD, puede consultar los siguientes recursos:

- *El Programa del Diploma: de los principios a la práctica*
- *Procedimientos de evaluación del Programa del Diploma*
- *Normas para los Colegios del Mundo del IB*
- Guías de las asignaturas (véanse las páginas de las asignaturas en el Centro de Recursos para los Programas)
- Materiales de ayuda al profesor (véanse las páginas de las asignaturas en el Centro de Recursos para los Programas)
- *La enseñanza y el aprendizaje guiados por la evaluación en el Programa del Diploma*

Programa de Orientación Profesional

Las características distintivas de la evaluación del Programa de Orientación Profesional (POP) son:

- El alumnado debe cumplir una serie de requisitos para conseguir el certificado del POP.
- No existe una puntuación total para el certificado del POP.
- Cuando el alumnado del POP estudia cursos junto con el alumnado del Programa del Diploma (PD), se realiza una evaluación común.
- El marco del POP exige que el alumnado realice unos estudios de formación profesional que el IB no ofrece y no evalúa.

Objetivos generales del Programa de Orientación Profesional

Los resultados de la evaluación solo son válidos si están claramente relacionados con los objetivos generales del curso y del programa.

La particularidad del POP es que prepara al alumnado para ingresar en el mundo laboral. El programa asegura que el alumnado desarrolle habilidades transferibles que le servirán para toda la vida en los estudios posteriores o el empleo para el que se quiera capacitar.

El programa permite al alumnado:

- Desarrollar una variedad amplia de competencias y profundizar su comprensión en áreas de conocimiento específicas mediante los cursos del PD y los estudios de formación profesional
- Desarrollar estrategias flexibles para la adquisición y profundización de conocimientos y habilidades en diversos contextos
- Fomentar actitudes y hábitos que le permitan mantener una actitud de aprendizaje durante toda la vida, con la voluntad de considerar nuevas perspectivas, oportunidades y desafíos
- Prepararse para una participación eficaz en un mundo cambiante
- Participar en un aprendizaje que desarrolle su capacidad y voluntad de contribuir de maneras constructivas

Usos válidos de los resultados de las evaluaciones del POP

Cuando el IB desarrolla los modelos de evaluación y los currículos, procura que las calificaciones finales obtenidas en los cursos y en el certificado del POP se puedan utilizar para determinar los siguientes aspectos:

- La selección para oportunidades laborales o programas de empleo, como el aprendizaje profesional
- La selección para estudios posteriores en el área de formación profesional correspondiente
- La selección para el ingreso a la universidad
- Si el alumno/a ya ha cumplido algunos requisitos de algún programa universitario (si se le pueden otorgar créditos o validar alguna asignatura o curso)

En los casos en los que el alumno/a ha realizado evaluaciones en una lengua (de respuesta) concreta, eso también demuestra que puede cursar estudios posteriores de esa disciplina o profesión en dicha lengua.

Para obtener más información sobre la importancia de estos usos válidos, consulte la sección “Validez” de esta publicación.

La estructura del POP

El POP es un marco educativo compuesto por tres partes, que son:

- Al menos dos cursos del PD de NM o NS
- Los componentes troncales del POP
- Estudios de formación profesional

Figura 43

Modelo del Programa de Orientación Profesional



Componentes troncales del POP

Los cuatro componentes troncales del POP están interrelacionados en lo que respecta a su naturaleza, sus resultados del aprendizaje y sus conexiones deliberadas. En su conjunto, tienen por objeto:

- Afianzar el programa sobre los principios del IB mediante el desarrollo de los atributos del perfil de la comunidad de aprendizaje del IB y la mentalidad internacional
- Contextualizar y mejorar los cursos del PD y los estudios de formación profesional a partir de todos los aspectos del marco
- Fomentar el desarrollo de conocimientos, habilidades y actitudes personales, académicos y profesionales duraderos
- Reflejar la convicción de que el aprendizaje es iterativo, está interconectado y es la piedra angular del bienestar personal y comunitario continuo

Es obligatorio completar los cuatro componentes troncales del POP, que son:

- Compromiso Comunitario
- Estudios Lingüísticos y Culturales
- Habilidades Personales y Profesionales
- Proyecto de Reflexión

El Proyecto de Reflexión se evalúa; esto significa que el proyecto y la reflexión finales reciben puntuaciones y el IB otorga al alumno/a una calificación final para este componente. Con respecto a los demás componentes troncales, el colegio debe confirmar al IB que el alumno/a los ha completado satisfactoriamente, pero el IB no comprueba la evaluación llevada a cabo por el colegio. Los procesos y prácticas empleados por los colegios que imparten el POP se revisan durante las visitas de evaluación del programa.

Cursos del PD como parte del POP

Todo el alumnado tiene que estudiar como mínimo dos cursos del PD. Al alumnado del POP y al del PD se les aplica el mismo proceso de evaluación. Es decir, ningún examen y ningún proceso de concesión de calificaciones finales es específico para el alumnado del POP.

A pesar de que hay cursos que son comunes para el PD y el POP, no está permitido cursar al mismo tiempo ambos programas. El alcance de los demás requisitos de cada programa impide que se puedan cursar simultáneamente.

Estudios de formación profesional

El IB no evalúa ni controla la calidad de los resultados de los estudios de formación profesional del POP. El único requisito es que el colegio confirme que el alumno/a ha completado los estudios. Los cursos del PD del IB no se prestan a integrarse en los estudios de formación profesional del POP.

Modo de calcular el resultado del POP

No existe una puntuación para el certificado del POP.

Si el alumno/a cumple todos los requisitos que se detallan a continuación, se le concederá el certificado del POP.

- El colegio ha confirmado que el alumno/a ha completado los estudios de formación profesional específicos.
- Haber obtenido, como mínimo, una calificación de 3 en al menos dos cursos del PD.
- Haber obtenido, como mínimo, una D en el Proyecto de Reflexión.
- El colegio ha confirmado que el alumno/a ha cumplido todos los requisitos de Compromiso Comunitario, Estudios Lingüísticos y Culturales, y Habilidades Personales y Profesionales.
- No haber recibido una sanción por conducta impropia del Comité de la Evaluación Final.

Las calificaciones y los resultados en el POP los confirma el mismo Comité de la Evaluación Final que para el PD.

Certificado bilingüe del POP

En lugar del certificado del POP habitual, un alumno/a puede obtener un certificado bilingüe del POP si cumple **al menos uno** de los siguientes criterios:

- Haber completado dos cursos de lengua de Estudios de Lengua y Literatura del PD, con una calificación mínima de 3 en cada uno de ellos.
- Haber completado un curso de lengua de Estudios de Lengua y Literatura del PD y también un curso de Individuos y Sociedades o de Ciencias del PD en una lengua de respuesta que no sea la misma que la del curso de Estudios de Lengua y Literatura. El alumno/a debe haber obtenido como mínimo una calificación de 3 en ambos cursos.

Gestión de los estándares y del trabajo faltante

Para el Proyecto de Reflexión y todos los cursos comunes con el PD, el POP utiliza los mismos enfoques que el PD en lo relativo al procedimiento de ausencia de puntuación y a la manera de gestionar los estándares según el tamaño del grupo de estudiantes.

Para obtener más información sobre el trabajo faltante y los casos de ausencia de puntuación, consulte la sección “Casos de ausencia de puntuación” de esta publicación.

Lecturas adicionales

Para obtener más información sobre el POP, puede consultar los siguientes recursos:

- *El Programa de Orientación Profesional: de los principios a la práctica*
- *Procedimientos de evaluación del Programa de Orientación Profesional*
- *Información general sobre el Programa de Orientación Profesional*
- *Guía del Proyecto de Reflexión*
- *Guía de Estudios Lingüísticos y Culturales*
- *Guía de Habilidades Personales y Profesionales*
- *Guía de Compromiso Comunitario*

Programa de los Años Intermedios

El certificado del PAI es un documento que reconoce que el alumno/a ha cumplido todos los aspectos del Programa de los Años Intermedios (PAI). Para obtener el certificado del PAI, es necesario:

- Sumar una puntuación total de al menos 28 puntos y no haber recibido ninguna calificación de 1 o 2 en ninguna evaluación electrónica del PAI
- Cursar el programa, preferiblemente, al menos durante los dos años recomendados (o durante un año como mínimo) y cumplir con los requisitos correspondientes al quinto año
- Completar los componentes de evaluación interna y los exámenes de un mínimo de cinco disciplinas de diferentes grupos de asignaturas y un máximo de ocho disciplinas de los ocho grupos de asignaturas, entre las que se deben incluir las asignaturas obligatorias
- Completar al menos una carpeta electrónica de los grupos de asignaturas de Artes, Educación Física y para la Salud o Diseño
- Completar el examen interdisciplinario en pantalla
- Completar y presentar un Proyecto Personal
- Completar los requisitos mínimos de Compromiso Comunitario del IB

Objetivos generales del Programa de los Años Intermedios

Los resultados de la evaluación solo son válidos si están claramente relacionados con los objetivos generales del curso y del programa.

El PAI fue concebido como un marco curricular coherente y completo que plantea un desafío académico y contribuye al desarrollo de habilidades valiosas para la vida en los alumnos de entre 11 y 16 años de edad, período fundamental en el desarrollo de los jóvenes. El éxito escolar está estrechamente relacionado con el bienestar personal, social y emocional. En una etapa en que los alumnos están estableciendo su identidad y consolidando su autoestima, el PAI puede motivarlos y ayudarlos a lograr los mejores resultados en el colegio y en la vida fuera del aula. Asimismo, les permite continuar desarrollando sus capacidades y enfrentar desafíos en asignaturas en las que quizá no se destaquen. El PAI ofrece a los alumnos la oportunidad de desarrollar su potencial, explorar sus propias preferencias de aprendizaje, afrontar riesgos adecuados y desarrollar una sólida noción de identidad personal y reflexionar sobre ella.

(El Programa de los Años Intermedios: de los principios a la práctica, 2014, p. 3).

En el PAI, los objetivos específicos de los grupos de asignaturas se corresponden claramente con los criterios de corrección. Todos los grupos de asignaturas del PAI tienen cuatro criterios de evaluación que coinciden con los cuatro objetivos específicos del grupo. Todos los criterios contribuyen de manera equivalente al resultado final.

Inclusión del contexto global en las evaluaciones electrónicas

En el PAI, los contextos de aprendizaje deben ser entornos, acontecimientos y circunstancias mundiales auténticos, o deben inspirarse en ellos. Los contextos para el aprendizaje se seleccionan entre contextos globales que estimulan la mentalidad internacional y el compromiso global dentro del programa. [...] En el PAI se establecen seis contextos globales para el aprendizaje y la enseñanza que se desarrollan a partir de los temas transdisciplinarios del PEP y los amplían.

(El Programa de los Años Intermedios: de los principios a la práctica, 2014, pp. 20-21).

Cada convocatoria de exámenes tendrá un contexto global específico y una exploración elegidos de la lista publicada en el documento *El Programa de los Años Intermedios: de los principios a la práctica*, en los cuales se basará.

Aproximadamente un tercio de las tareas del examen en pantalla de cada disciplina se relacionarán con el contexto global seleccionado, o bien se inspirarán o basarán en dicho contexto. La totalidad del examen interdisciplinario en pantalla se inspira en el contexto global seleccionado.

Se elaborarán planificadores de unidades parcialmente completados de Adquisición de Lenguas, Artes, Diseño, y Educación Física y para la Salud, que harán referencia al contexto global elegido.

Usos válidos de los resultados de las evaluaciones del PAI

Cuando el IB desarrolla los modelos de evaluación y los currículos, procura que las calificaciones finales obtenidas en los cursos del PAI y la puntuación obtenida en el certificado se puedan utilizar para determinar los siguientes aspectos:

- La selección para estudios posteriores o para el mundo laboral
- Comentarios positivos e indicaciones de puntos fuertes del alumnado que continúa con sus estudios

En los casos en los que el alumno/a ha realizado la evaluación en una lengua (de respuesta) concreta, eso también demuestra que puede cursar estudios posteriores en dicha lengua (por ejemplo, estudiar en un colegio de habla española, francesa o inglesa).

Para obtener más información sobre la importancia de estos usos válidos, consulte la sección “Validez” de esta publicación.

La estructura del PAI

Figura 44

Modelo del Programa de los Años Intermedios



Modo de calcular el resultado del PAI

Al final del quinto año del PAI, el alumnado puede inscribirse en la evaluación externa del IB. Los resultados de estas evaluaciones quedarán registrados en el documento *resultados de los cursos del PAI*. Además, este alumnado puede decidir si quiere realizar las evaluaciones con las que se le podría conceder el certificado del PAI.

El colegio también puede emitir un título de participación del PAI, dirigido al alumnado que estudia el programa durante un mínimo de dos años y cumple los requisitos en los años tercero o cuarto. Estos alumnos/as no están matriculados para ningún tipo de evaluación con el IB. El título de participación del PAI es un documento del colegio y no está verificado por el IB.

Para conseguir el certificado del PAI, el alumnado debe haber cursado el último año del programa, con un período de participación recomendado de dos años, y también debe:

- Completar una evaluación en pantalla o bien una carpeta electrónica en seis asignaturas elegidas de los siguientes grupos: Lengua y Literatura, Adquisición de Lenguas (o una segunda asignatura de Lengua y Literatura), Individuos y Sociedades, Matemáticas, Ciencias y, por último, una asignatura de Artes, Educación Física y para la Salud, o Diseño
- Obtener una calificación de 3, como mínimo, en cada una de las seis asignaturas mencionadas en el punto anterior
- Completar el examen interdisciplinario en pantalla y obtener una calificación de 3, como mínimo
- Completar el Proyecto Personal y obtener una calificación de 3, como mínimo
- Obtener una puntuación total de 28 puntos, como mínimo
- Cumplir las expectativas de Compromiso Comunitario del colegio

El certificado bilingüe del PAI requiere además obtener resultados satisfactorios en los exámenes en pantalla de uno de los siguientes elementos:

- Un segundo curso de Lengua y Literatura (en lugar de un curso de Adquisición de Lenguas)
- Uno (o varios) cursos de Ciencias o Individuos y Sociedades, o el examen interdisciplinario, en una lengua distinta de la elegida para el curso de Lengua y Literatura

La evaluación sumativa externa (evaluación electrónica del PAI)

La evaluación electrónica opcional contempla dos formas de determinar los conocimientos y las competencias del alumnado:

- Carpetas electrónicas que contienen trabajos entregados para evaluación de Adquisición de Lenguas, Artes, Diseño, y Educación Física y para la Salud, que posteriormente se someten a moderación para garantizar que se sigue un estándar global coherente
- A través de exámenes en pantalla (dos horas de duración) para determinados cursos de Lengua y Literatura, Adquisición de Lenguas, Individuos y Sociedades, Ciencias, Matemáticas y aprendizaje interdisciplinario

Además, el Proyecto Personal se entrega electrónicamente al IB para su moderación. Si bien hay otros componentes de la evaluación electrónica que son opcionales para los colegios, todo el alumnado del quinto año del PAI debe realizar el Proyecto Personal.

Modelos de exámenes

El IB publica modelos de exámenes para que los colegios tengan una orientación clara sobre cómo serán las evaluaciones electrónicas. Estos modelos permiten a docentes y estudiantes comprender la naturaleza y el propósito de la evaluación electrónica del PAI. Asimismo, ayudan al alumnado a prepararse para los exámenes en pantalla y centrar la atención en los criterios de los grupos de asignaturas y las estrategias de

evaluación de cada grupo. En todos los modelos siempre hay cuatro criterios, cada uno de los cuales tiene el mismo valor que los demás con respecto al total de la evaluación.

El IB se asegura de que, en todas las convocatorias, los exámenes no se desvíen de los modelos en más de 3 puntos.

Carpetas electrónicas y planificadores de unidades parcialmente completados

Las carpetas electrónicas permiten evaluar el desempeño en representaciones o trabajos de clase (productos) que se realizan durante un largo período de tiempo, y cuya evaluación mediante un examen resultaría muy difícil debido a su naturaleza. Las carpetas electrónicas se basan en planificadores de unidades parcialmente completados que son lo suficientemente flexibles como para adaptarse a los contextos locales, y que ayudan al profesorado a asegurarse de que los trabajos del alumnado proporcionen las pruebas adecuadas que permitan emitir juicios justos y pertinentes. En cada convocatoria se proporcionan nuevos planificadores de unidades parcialmente completados.

Los planificadores de unidades deben procurar que las tareas que encargue el profesorado permitan al alumnado alcanzar todas las calificaciones del PAI. Las evaluaciones que diseña el profesorado implican el riesgo de poner al alumnado en desventaja si las tareas son más fáciles o más difíciles de lo que es razonable. En la moderación, el IB solamente puede conceder calificaciones basándose en el trabajo entregado para evaluación. Si la tarea que diseña el profesor/a solamente corresponde a una parte del rango de descriptores de calificaciones finales, no se podrá conceder ninguna calificación fuera de dicho rango.

La evaluación única y la gestión de la carga de trabajo del alumnado

Para el alumnado, realizar los exámenes y el trabajo de clase supone someterse a cierto grado de estrés y exigencia. Además, tanto los exámenes como el trabajo de clase pueden restar tiempo al destinado a la enseñanza. El IB considera conveniente reducir al mínimo la cantidad de evaluaciones sumativas para el alumnado del quinto año del PAI. Si bien esto provoca dificultades porque solo tienen una oportunidad para demostrar sus logros, el IB las acepta para procurar el bienestar general de cada estudiante.

Método para determinar las calificaciones ausentes

Uno de los inconvenientes de querer reducir la carga de trabajo del alumnado es que, al utilizar solo un componente de evaluación, las pruebas disponibles en el PAI son escasas. Esto significa que el procedimiento de ausencia de calificación del IB es menos fiable que el procedimiento de ausencia de puntuación en los otros programas, por lo cual solo se debe utilizar en casos excepcionales.

Para obtener información detallada sobre el proceso de ausencia de calificación, consulte la sección “Procedimiento de ausencia de calificación” de esta publicación.

Lecturas adicionales

Para obtener más información sobre el PAI, puede consultar los siguientes recursos:

- [*El Programa de los Años Intermedios: de los principios a la práctica*](#)
- [*Guía para la convocatoria de exámenes del PAI*](#)
- [*Guías de las asignaturas del PAI*](#)
- [*Guías de los proyectos del PAI*](#)
- [*Procedimientos de evaluación del Programa de los Años Intermedios*](#)
- [*Requisitos de TI y responsabilidades del colegio para los exámenes en pantalla del PAI*](#)
- [*Familiarización con los exámenes en pantalla del PAI destinada a los alumnos*](#)

Programa de la Escuela Primaria

Las características distintivas de la evaluación del Programa de la Escuela Primaria (PEP) son:

- La evaluación implica la colaboración de docentes y alumnado para el seguimiento, la documentación, la medición, la comunicación de resultados y la adaptación del aprendizaje.
- La promoción de una cultura de evaluación supone el desarrollo de la capacidad de evaluación de cada miembro de la comunidad de aprendizaje.
- El IB no impone ninguna evaluación sumativa externa.

Objetivos generales del Programa de la Escuela Primaria

Los resultados de la evaluación solo son válidos si están claramente relacionados con los objetivos generales del curso y del programa.

El marco curricular del PEP se centra en el aprendizaje transdisciplinario como organizador del currículo con el objetivo de que el alumnado experimente un aprendizaje que se ubique dentro de los límites disciplinarios tradicionales y que los trascienda.

Inspirado en teorías del aprendizaje constructivistas y socioconstructivistas, el marco curricular del PEP comienza con la premisa de que el alumnado del programa es agente de su propio aprendizaje y colaborador en el proceso de aprendizaje. El alumnado posee un potencial innato para indagar, cuestionar, hacerse preguntas y teorizar acerca de sí mismo, de otras personas y del mundo que les rodea.

Esta noción de cómo aprende el alumnado conforma las bases de un modelo de aprendizaje y enseñanza transdisciplinario, basado en la indagación e impulsado por conceptos. Al participar en el programa de indagación y reflexionar sobre su aprendizaje, el alumnado del PEP desarrolla conocimientos, habilidades y actitudes para desenvolverse en el mundo, mediante la realización de acciones en favor del bienestar de las personas y del planeta.

La estructura del PEP

Figura 45

Modelo del Programa de la Escuela Primaria



La evaluación en el PEP

En el PEP, la evaluación es una parte central del objetivo de apoyar al alumnado cuidadosa y eficazmente a través de la adquisición de conocimientos y habilidades específicas de las asignaturas, la comprensión de conceptos, y el desarrollo de las habilidades de los enfoques del aprendizaje y de los atributos del perfil de la comunidad de aprendizaje del IB.

La evaluación del PEP tiene cuatro dimensiones: el seguimiento, la documentación, la medición y la comunicación de los resultados del aprendizaje. Cada uno de estos aspectos desempeña su propia función, pero todos tienen como objetivo proporcionar indicios que sirvan de base para el aprendizaje y la enseñanza. Si bien no todas las dimensiones de la evaluación cuentan con la misma ponderación, cada una de ellas tiene su propia importancia y valor. En el PEP se hace hincapié en el seguimiento y la documentación del aprendizaje, ya que estas dimensiones son fundamentales para proporcionar al alumnado comentarios que pueden ponerse en práctica.

Figura 46

Las dimensiones de la evaluación en el PEP



Los siguientes supuestos sirven de base a la evaluación en el PEP:

- La evaluación es un proceso continuo y colaborativo entre docente y estudiante, y entre compañeros/as, que consta de la recopilación, análisis, reflexión y acción con respecto a las pruebas del aprendizaje, que sirve para orientar la enseñanza.
- Para evaluar el trabajo, se debe emplear una variedad de estrategias que tomen en cuenta las diversas, complejas y sofisticadas maneras en que cada alumno/a comprende sus experiencias.
- El alumnado participa activamente en la evaluación y reflexión acerca de su aprendizaje, teniendo en cuenta los comentarios de sus compañeros/as y docentes para fomentar la proactividad en los pasos siguientes del aprendizaje.
- Los objetivos de aprendizaje y los criterios de logro se construyen de manera conjunta y se comunican claramente.
- Tanto los resultados del aprendizaje como el proceso de aprendizaje son objeto de evaluación.
- El diseño de la evaluación incluye una planificación inversa y prospectiva.
- La promoción de una cultura de evaluación supone el desarrollo de la capacidad de evaluación de cada miembro de la comunidad de aprendizaje.

Lecturas adicionales

Para obtener más información sobre el PEP, consulte la publicación *El Programa de la Escuela Primaria: de los principios a la práctica* en el Centro de Recursos para los Programas.

La moderación de la evaluación interna en mayor detalle

La moderación se utiliza en los trabajos de evaluación interna del Programa de los Años Intermedios (PAI), el Programa del Diploma (PD) y el Programa de Orientación Profesional (POP) para garantizar que se aplique un estándar común en todos los colegios. Como resultado de la moderación, las puntuaciones de un colegio pueden subir, bajar o quedarse igual. El objetivo de la moderación es comprobar con qué precisión y coherencia ha aplicado el profesor/a los criterios de evaluación al corregir el trabajo del alumnado.

Muestras

Una vez que el colegio ha enviado las puntuaciones de la evaluación interna, el Bachillerato Internacional (IB) selecciona al alumnado que formará parte de la muestra para la moderación. Esta muestra se selecciona cuidadosamente para garantizar que el rango de puntuaciones del colegio esté representado adecuadamente. El número de trabajos que componen la muestra enviada para la moderación puede ser de diez, ocho, cinco o menos de cinco, dependiendo de la cantidad de alumnos/as matriculados en la asignatura. Dentro de la muestra, se asignarán tres alumnos/as a la muestra inicial, para representar el rango de puntuaciones de la muestra y determinar si las puntuaciones del profesor/a están dentro de los márgenes de tolerancia del estándar global. Si no es así, el examinador/a también corregirá el resto de la muestra (que pasará a denominarse *muestra ampliada*). Para la muestra que se somete a moderación, el IB no suele seleccionar a estudiantes que hayan logrado la máxima puntuación, ya que así permite que el alumnado que esté dentro del rango de puntuaciones más elevado tenga la posibilidad de que sus puntuaciones suban a raíz de la moderación.

Cuando en un colegio hay suficientes alumnos/as matriculados en un curso determinado como para que puedan repartirse en más de una clase, y más de un profesor/a realiza la evaluación interna, el IB espera que los profesores/as compartan la evaluación interna y colaboren para estandarizar el modo de aplicar los criterios. Al colegio se le pide una única muestra para la moderación, que con toda probabilidad contendrá trabajos entregados para evaluación corregidos por los distintos profesores/as. No obstante, cuando haya en un colegio distintas clases que utilicen diferentes lenguas de respuesta para la misma asignatura, se requiere una muestra distinta para cada lengua.

El cálculo del factor de moderación

Todos los componentes de evaluación interna se corrigen aplicando criterios de evaluación o bandas de puntuación. La mayoría de veces, el profesor/a tiene acceso a más cantidad de información que el examinador/a sobre el contexto y el proceso que rodea al trabajo entregado para evaluación. Por ese motivo, a los examinadores/as que moderan los componentes de evaluación interna se les pide que juzguen si la corrección del profesor/a resulta adecuada; no se trata de volver a corregir el trabajo y rechazar las puntuaciones del profesor/a. De hecho, esas puntuaciones solamente se pueden modificar si el moderador/a está seguro de que resultan inadecuadas y no se corresponden con el estándar de corrección global.

El examinador/a modera la muestra de trabajos y, si es necesario, hace un ajuste a las puntuaciones del profesor/a para todo el alumnado del colegio en ese componente; para ello, se basa en una comparación estadística entre los dos grupos de puntuaciones.

Si las puntuaciones que otorga el profesor/a son sistemáticamente demasiado bajas o demasiado altas, se realizará el mismo ajuste en todas las puntuaciones que otorgue.

En el siguiente ejemplo (tabla 8), las puntuaciones que ha concedido el profesor/a han sido, en promedio, 5 puntos demasiado bajas. Por tanto, en ese grupo de estudiantes se añadirán 5 puntos a todas las puntuaciones del profesor/a.

Tabla 8
Puntuaciones sistemáticamente demasiado bajas

Puntuación del profesor	Puntuación del examinador	Diferencia entre las puntuaciones del profesor y el examinador	Puntuación final moderada
15	20	5	20
11	17	6	16
10	14	4	15
8	12	4	13
4	10	6	9

Si las puntuaciones que otorga el profesor/a son demasiado bajas o demasiado altas únicamente en la parte superior o en la parte inferior del rango de puntuaciones, es posible que el ajuste varíe en todo el rango de puntuaciones.

En el siguiente ejemplo (tabla 9), el profesor/a ha corregido de forma incoherente en comparación con el estándar global. Las puntuaciones han sido demasiado generosas en la parte superior del rango, pero demasiado severas en la parte inferior. Por tanto, el ajuste que se aplique a las puntuaciones del profesor/a variará en todo el rango.

Tabla 9
Corrección incoherente

Puntuación del profesor	Puntuación del examinador	Diferencia entre las puntuaciones del profesor y el examinador	Puntuación final moderada
37	35	-2	35
25	22	-3	24
17	18	1	17
12	14	2	13
6	8	2	8

Si el profesor/a está empleando el estándar correcto en su corrección, no se realizará ningún ajuste.

En la tabla 10, todas las puntuaciones que ha concedido el profesor/a están dentro de los márgenes de tolerancia del estándar global. Por tanto, todo el alumnado recibirá las puntuaciones del profesor/a.

Tabla 10
Corrección de acuerdo con el estándar

Puntuación del profesor	Puntuación del examinador	Diferencia entre las puntuaciones del profesor y el examinador	Puntuación final moderada
32	31	-1	32
29	30	1	29
8	8	0	8

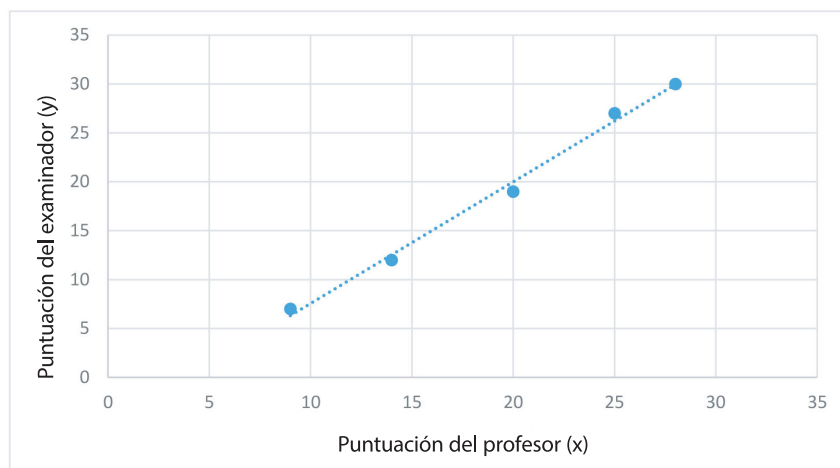
Regresión lineal

En cada muestra para la moderación se analizan los datos, lo cual permite determinar de manera inicial si la corrección de un profesor/a está dentro de los márgenes de tolerancia. Si no lo está, se aplicará un ajuste adecuado a todas las puntuaciones del profesor/a en función de la tendencia general que se vea en la muestra. Para ello se utiliza la *regresión lineal*, que implica calcular la línea recta que mejor se ajuste al conjunto de puntos de datos correspondientes a las puntuaciones que el profesor/a y el examinador/a han concedido a los trabajos de la muestra. En la figura 47 se muestra un ejemplo de regresión lineal y su correspondiente conjunto de puntuaciones.

En la figura 47 se muestra una recta de regresión de la moderación para un profesor/a cuyas puntuaciones son levemente severas en el extremo superior del rango y demasiado generosas en el inferior. Cada uno de los puntos representa las dos puntuaciones concedidas a un trabajo de muestra: una del profesor/a y la otra del examinador/a. La recta de regresión continua se utiliza para convertir las puntuaciones del profesor/a en puntuaciones moderadas.

Figura 47

Regresión lineal y su correspondiente conjunto de puntuaciones



Puntuación del profesor	Puntuación del examinador	Puntuación final moderada
28	30	30
25	27	26
20	19	20
14	12	12
9	7	6

La ecuación de la recta de regresión, que se calcula a partir de los datos de la muestra, se puede utilizar para convertir cada puntuación (x) otorgada por el profesor/a en una puntuación (y) equivalente que el examinador/a probablemente hubiera concedido a ese mismo alumno/a. Este ajuste de moderación, basado en extrapolar una muestra a un conjunto mucho más amplio de puntuaciones, solo puede reflejar las tendencias generales que se observan en la corrección. El propósito de la moderación es garantizar que, en conjunto, las puntuaciones del alumnado se ajustan a niveles más adecuados. Se trata de un enfoque en el que se busca el ajuste más adecuado, por lo que no todo el alumnado recibirá la puntuación del examinador/a como puntuación moderada.

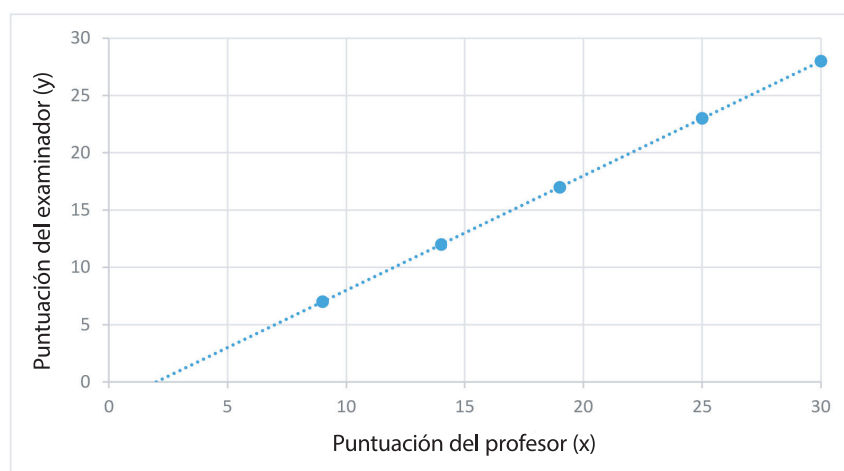
Error en la moderación

Automáticamente se realiza una comprobación para asegurarse de que la recta de regresión lineal (es decir, el factor de moderación calculado) cumple determinadas condiciones antes de aplicarse a todas las puntuaciones de un profesor/a. En algunos casos, tal vez no sea posible calcular un ajuste de moderación con el trabajo de muestra entregado. Una medida estadística es el coeficiente de correlación (se utiliza el coeficiente de correlación producto-momento), que mide la coherencia de la relación entre las puntuaciones del profesor/a y las del examinador/a.

Un coeficiente de correlación de 1 indica que hay una coherencia perfecta en la relación entre las puntuaciones, y que hay coincidencia en la clasificación del alumnado de mejor a peor (aunque no necesariamente las mismas puntuaciones), como se muestra en la figura 48.

Figura 48

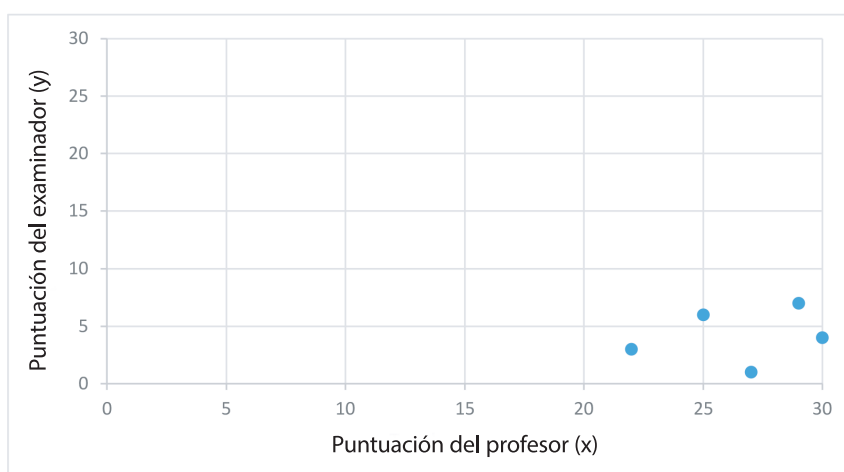
Coeficiente de correlación de 1



Un coeficiente de 0 indica que no hay ninguna relación, como se muestra en la figura 49.

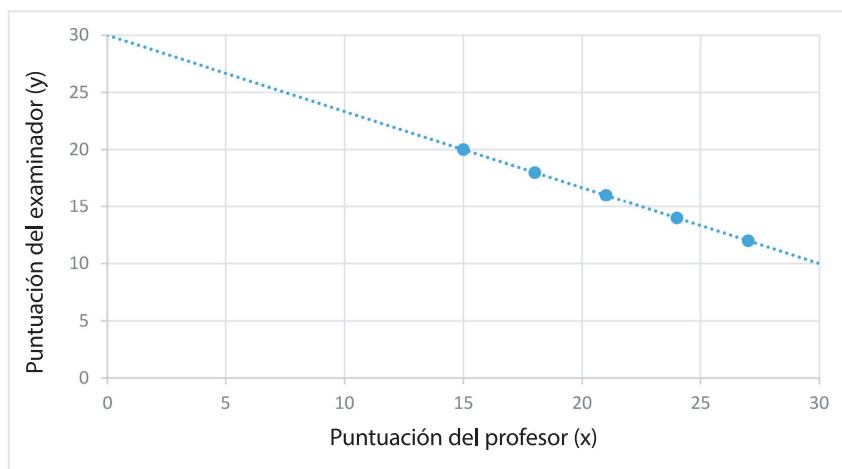
Figura 49

Coeficiente de correlación de 0



Un coeficiente de -1 indica que hay un desacuerdo constante en relación con los méritos relativos del trabajo del alumnado, y que el profesor/a y el examinador/a producen clasificaciones opuestas, como se muestra en la figura 50.

Figura 50
Coeficiente de correlación de -1



Para que el factor de moderación sea aceptable, el coeficiente de correlación debe ser de al menos 0,85, lo que indica un alto nivel de coincidencia entre el profesor/a y el examinador/a. Sin embargo, un coeficiente de correlación alto no basta por sí solo para garantizar que el factor de moderación resulta adecuado. Por tanto, se comprueba que el gradiente (la pendiente) de la recta de regresión esté entre 0,5 y 1,5. Si el gradiente de la recta es demasiado bajo, significa que el profesor/a ha utilizado un rango de puntuaciones demasiado grande, y en comparación ha concedido muy pocos puntos al trabajo de menor calidad y demasiados puntos al trabajo de mayor calidad, aunque puede haberlo hecho de manera coherente. En consecuencia, el examinador/a ha tenido que comprimir considerablemente el rango de puntuaciones del profesor/a. Si el gradiente es superior a 1,5, la recta es demasiado inclinada y sucede lo contrario: el profesor/a no ha diferenciado lo suficiente entre el trabajo entregado de menor y de mayor calidad, y el examinador/a ha tenido que ampliar el rango de puntuaciones concedido.

Una muestra no superará las comprobaciones automáticas de la moderación si el coeficiente de correlación es inferior a 0,85 o si el gradiente no está entre 0,5 y 1,5.

El personal de evaluación del IB revisa individualmente todos los casos de muestras de colegios que no superan la moderación; para ello, tiene en cuenta los datos subyacentes y puede decidir que:

- La recta de regresión calculada es adecuada para el rango de puntuaciones concreto del profesor/a.
- Hay que aplicar otro ajuste de moderación que resulte adecuado para el rango de puntuaciones del profesor/a.
- Hay que solicitar más datos de muestras para poder aclarar la tendencia.
- Hay que solicitar los trabajos del alumnado restante para poder realizar una revisión completa de la corrección del profesor/a.

Todos los trabajos de evaluación interna del alumnado deben estar disponibles hasta que finalice la convocatoria de exámenes pertinente, con el fin de que el IB pueda resolver cualquier error en la moderación.

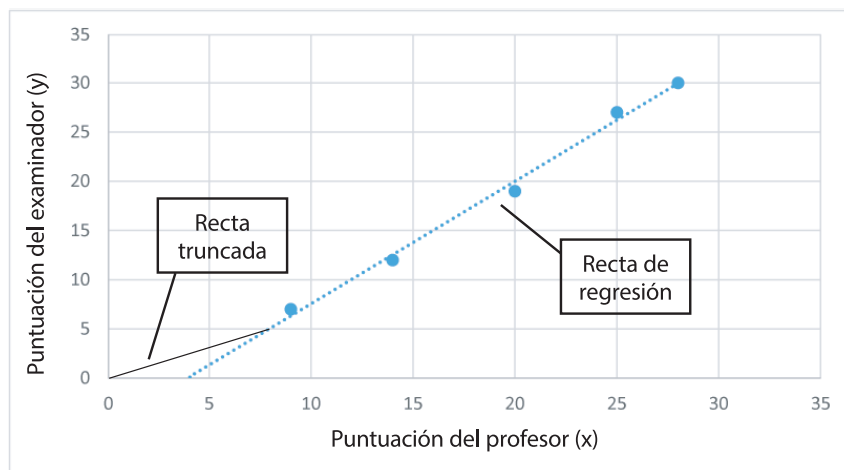
Adaptación del modelo lineal

El modelo de la recta empleado para la moderación se modifica en cierto modo utilizando *truncamientos*. El ajuste de moderación obtenido mediante una línea recta puede producir efectos inapropiados en los extremos del rango de puntuaciones posibles, de manera que resulta imposible que un alumno/a obtenga un 0. Es posible que a un alumno/a cuyo trabajo merece realmente un 0 se le concedan algunos puntos incluso aunque no haya escrito nada que valga la pena en su trabajo, o puede que a un alumno/a cuyo trabajo es de poca calidad, pero aun así merezca algunos puntos, reciba un 0 después de la moderación.

Para superar este problema, se puede aplicar el truncamiento a las puntuaciones que están en el 20 % inferior del rango de puntuaciones disponibles. En ese extremo, la recta de regresión calculada se modifica y la sustituye una nueva recta truncada que enlaza de la recta de regresión a las coordenadas mínimas. El truncamiento de la recta de regresión no se aplica en el extremo superior, ya que a menudo hay docentes que conceden la puntuación máxima a estudiantes que no la merecen.

En el gráfico de la figura 51 se muestra el truncamiento de una recta de regresión para evitar que las puntuaciones del alumnado se ajusten al acercarse mucho a los valores mínimos.

Figura 51
Recta de regresión truncada



El truncamiento garantiza que una puntuación moderada de 0 solo puede surgir de una puntuación que originalmente sea 0. También evita que se le dé una puntuación de 0 a los trabajos que merecen una puntuación baja, y que el trabajo sin valor alguno reciba una puntuación baja. Para esto, se da por hecho que la corrección del profesor/a supera el proceso de moderación automática. Si la puntuación del profesor/a no supera la moderación y no se puede moderar automáticamente, no se aplica el truncamiento.

Bibliografía

- ANANIADOU, K. y CLARO, M. "21st century skills and competences for new millennium learners in OECD countries". En *OECD Education Working Papers*. 2009, n.º 41.
- ANDERSON, L. W.; KRATHWOHL, D. R. *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman, 2001.
- BACHILLERATO INTERNACIONAL. Declaración de principios del IB. Organización del Bachillerato Internacional, 2024.
- BACHILLERATO INTERNACIONAL. *¿Qué es la educación del IB?* Organización del Bachillerato Internacional, 2019.
- BAIRD, J.; CRESSWELL, M. J. y NEWTON, P. "Would the real gold standard please step forward?". En *Research Papers in Education*. 2000, vol. 15, n.º 2. Pp. 213-229. <<https://doi.org/10.1080/026715200402506>>.
- BIGGS, J. "Enhancing teaching through constructive alignment". En *Higher Education*. 1996, vol. 32, n.º 3. Pp. 347-364. <<https://www.jstor.org/stable/3448076>>.
- BLACK, P. "Assessment, Learning Theories and Testing Systems". En Murphy, P. (ed.). *Learners, learning & assessment*. Paul Chapman Publishing en asociación con The Open University. 1999. Pp. 118-134.
- BLOOM, B. S. (ed.); ENGLEHART, M. D.; FURST, E. J.; HILL, W. H. y KRATHWOHL, D. R. *Taxonomy of educational objectives: Handbook 1: Cognitive domain*. Longman, 1956.
- BROADFOOT, P. M. *Education, assessment and society: A sociological analysis*. Open University Press, 1996.
- BROWN, R. "Cultural dimensions of national and international educational assessment". En Hayden, M; Thompson, J. y Walker, G. (eds.). *International education in practice: Dimensions for schools and international schools*. Routledge, 2002.
- CHAMBERLAIN, S. "AQA: Public perceptions of reliability". En Ofqual. *Ofqual's Reliability Compendium*. Capítulo 18. The Office of Qualifications and Examinations Regulation, 2010.
- COE, R.; SEARLE, J.; BARMBY, P.; JONES, K. y HIGGINS, S. *Relative difficulty of examinations in different subjects*. [Informe para SCORE (Science Community Supporting Education)]. CEM Centre, Universidad de Durham, 2008.
- CRESSWELL, M. J. "Defining, setting and maintaining standards in curriculum-embedded examinations: Judgemental and statistical approaches". En Goldstein, H. y Lewis, T. (eds.). *Assessment: Problems, developments and statistical issues*. Wiley, 1996. Pp. 57-84.
- CRESSWELL, M. J. "Examination grades: How many should there be?". En *British Educational Research Journal*. 1986, vol. 12, n.º 1. Pp. 37-54.
- CRESSWELL, M. J. "The role of public examinations in defining and monitoring standards". En *Proceedings of the British Academy*. 2000, vol. 102. Pp. 69-120.
- CROOKS, T. J.; KANE, M. T. y COHEN, A. S. "Threats to the valid use of assessments". En *Assessment in Education: Principles, Policy & Practice*. 1996, vol. 3, n.º 3. Pp. 265-286. <<https://doi.org/10.1080/0969594960030302>>.
- DOLAN, R. P.; BURLING, K.; HARMS, M.; STRAIN-SEYMOUR, E.; WAY, W. y ROSE, D. *A universal design for learning-based framework for designing accessible technology-enhanced assessments*. (Informe de investigación). Pearson, 2013. <<https://eric.ed.gov/?id=ED576691>>.
- FRITH, D. S. y MACINTOSH, H. G. *A teacher's guide to assessment*. Stanley Thornes, 1984.
- GIBBS, G. *Teaching more students: Assessing more students*. The Oxford Centre for Staff Learning and Development, Universidad Oxford Brookes, 1992.
- GIPPS, C. y MURPHY, P. *A Fair Test? Assessment, Achievement and Equity*. Open University Press, 1994.

- GLASER, R. "Instructional technology and the measurement of learning outcomes: Some questions". En *American Psychologist*. 1963, vol. 18, n.º 8. Pp. 519-521. <<https://doi.org/10.1037/h0049294>>.
- GOLDSTEIN, H. "Group differences and bias in assessment". En Goldstein, H. y Lewis, T. (eds.). *Assessment: Problems, developments and statistical issues*. John Wiley & Sons Ltd. 1996. Pp. 85-93.
- GOOD, F. J. y CRESSWELL, M. J. *Grading the GCSE*. Secondary Examinations Council, 1988.
- HE, Q.; OPPOSS, D. y BOYLE, A. "A quantitative investigation into public perceptions of reliability in examination results in England". En Ofqual. *Ofqual's Reliability Compendium*. Capítulo 19. The Office of Qualifications and Examinations Regulation, 2010. Pág. 68.
- HIERONYMUS, A. N. y HOOVER, H. D. *Iowa tests of basic skills: Manual for school administrators*. Riverside Publishing Company, 1986.
- HUGHES, D. C.; KEELING, B. y TUCK, B. F. Effects of achievement expectations and handwriting quality on scoring essays. En *Journal of Educational Measurement*. 1983, vol. 20, n.º 1. Pp. 65-70. <<https://www.jstor.org/stable/1434941>>.
- HUMPHREYS, L. G. "An analysis and evaluation of test and item bias in the prediction context". En *Journal of Applied Psychology*. 1986, vol. 71, n.º 2. Pp. 327-333. <<https://doi.org/10.1037/0021-9010.71.2.327>>.
- LAMBERT, D. y LINES, D. *Understanding assessment: Purposes, perceptions, practice*. Routledge Falmer, 2000.
- LINN, M. C. "Gender differences in educational achievement". En Pfeleiderer, J. (ed.). *Sex equity in educational opportunity, achievement, and testing*. Educational Testing Service, 1992.
- LLEWELLYN, D. *Inquire within: Implementing inquiry- and argument-based science standards in grades 3-8*. 3.ª edición. Corwin, 2014.
- MEYER, A.; ROSE, D. H. y GORDON, D. *Universal Design for Learning: Theory and Practice*. CAST Professional Publishing, 2014.
- MITRA, S. "Responsibility, leadership and education". [Ponencia]. Conferencia para directores/as de Colegios del Mundo del IB, Singapur. 13-16 de octubre de 2011.
- MURPHY, P. *Learners, learning & assessment*. Paul Chapman Publishing en asociación con The Open University. 1999.
- NATIONAL RESEARCH COUNCIL. *Assessing 21st century skills: Summary of a workshop*. The National Academies Press, 2011.
- NEWTON, P. "A level pass rates and the enduring myth of norm-referencing". En *Research Matters*. 2011. <<https://doi.org/10.17863/CAM.100441>>.
- NEWTON, P. E. "Clarifying the purposes of educational assessment". En *Assessment in Education: Principles, Policy & Practice*. 2007, vol. 14, n.º 2. Pp. 149-170. <<https://doi.org/10.1080/09695940701478321>>.
- NEWTON, P. E. "We need to talk about validity". [Ponencia]. Reunión anual del National Council for Measurement in Education. Vancouver (Canadá), 2012.
- NEWTON, P. E. y SHAW, S. D. *Validity in educational and psychological assessment*. Londres (Reino Unido): SAGE Publications, 2014.
- PETERSON, A. D. C. *New techniques for the assessment of pupils' work*. Consejo de Europa, 1971.
- PETERSON, A. D. C. *Schools across frontiers: The story of the International Baccalaureate and the United World Colleges*. 2.ª edición. Open Court Publishing Company, 2003.
- POPHAM, W. J. *Criterion-referenced measurement*. Prentice Hall, 1978.
- RAND CORPORATION. "Teaching and learning 21st century skills: Lessons from the learning sciences". Asia Society, Abril de 2012. <http://www.rand.org/pubs/external_publications/EP51105.html>. [Consulta: 27-11-2025].
- RAO, K.; CURRIE-RUBIN, R. y LOGLI, C. *UDL and inclusive practices in IB schools worldwide*. CAST Professional Learning, 2016.
- SCHLEICHER, A. "The case for 21st-century learning". En OCDE. *OECD.org* [en línea]. 2016. <<https://web.archive.org/2012-06-14/61660-the-case-for-21st-century-learning.htm>>. [Consulta: 27-11-2025].

- SHEPARD, L. A. "Commentary: what policy makers who mandate tests should know about the new psychology of intellectual ability and learning". En Gifford, B. R. y O'Connor, M. C. (eds.). *Changing assessments: Alternative views of aptitude, achievement and instruction* Kluwer Academic Publishers, 1992. Pp. 301-328.
- SIRECI, S. G. "On test validity theory and test validation". En *Educational Researcher*. 2007, vol. 36, n.º 8. Pp. 477-481. <<https://www.jstor.org/stable/4621103>>.
- SMITH, D. J. y TOMLINSON, S. *The school effect: A study of multi-racial comprehensives*. Policy Studies Institute, 1989.
- SNYDER, B. R. *The Hidden Curriculum*. MIT Press, 1971.
- SURGENOR, P. *Teaching toolkit: Effect of assessment on learning*. UCD Dublin, 2010.
- VYGOTSKY, L. S. *Thought and language*. MIT Press, 1962.
- WILIAM, D. "Validity, dependability and reliability in National Curriculum assessment". En *The Curriculum Journal*. 1993, vol. 4, n.º 3. Pp. 335-350. <<https://doi.org/10.1080/0958517930040303>>.
- WINKLEY, J. y CRESSWELL, M. "Introduction to the concept of reliability". En Ofqual. *Ofqual's Reliability Compendium*. Capítulo 1. The Office of Qualifications and Examinations Regulation, 2011.
- WOOD, D.; BRUNER, J. S. y ROSS, G. "The role of tutoring in problem solving". En *Journal of Child Psychology and Psychiatry*. 1976, vol. 17, n.º 2. Pp. 89-100. <<https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>>.
- WOOD, D. *How children think and learn: The social contexts of cognitive development*. 2.ª edición. Blackwell Publishing, 1998.
- ZANGA, G. y DE GIOANNIS, E. "Discrimination in grading: A scoping review of studies on teachers' discrimination in school". En *Studies in Educational Evaluation*. 2023, vol. 78. <<https://doi.org/10.1016/j.stueduc.2023.101284>>.

Glosario

El IB está comprometido a utilizar un lenguaje sencillo y a reducir el uso de terminología específica. Para apoyar este compromiso, este glosario define algunas palabras y conceptos clave que aparecen en este y otros recursos relacionados con la evaluación. Algunos de estos términos también se pueden utilizar en las comunicaciones del IB.

Término	Definición
Acceso inclusivo	Evaluación que tiene en cuenta las necesidades de todo el alumnado para que pueda demostrar de manera justa su competencia en la asignatura.
Adecuaciones inclusivas de acceso	Condiciones adicionales o modificadas durante el proceso de evaluación para un alumno/a con necesidades específicas de acceso a la evaluación. Estas permiten al alumno/a demostrar su nivel de logro de manera más justa.
Archivo de respuestas del alumno	En los exámenes en pantalla del PAI, este término se refiere al trabajo que completa un alumno/a durante el examen. Dicho trabajo puede incluir contenido multimedia. El proceso de creación de la respuesta forma parte de lo que se evalúa. La respuesta se entrega como archivo electrónico, que a menudo se denomina <i>archivo de respuestas del alumno/a</i> .
Argumento de validez	Pruebas y explicaciones que respaldan las decisiones tomadas al crear una evaluación y que justifican que esta es adecuada al propósito.
Asignatura piloto	Asignatura en fase de prueba. Si el resultado de la prueba es satisfactorio, la asignatura piloto se ofrecerá de manera general.
Banco de preguntas	Recopilación de preguntas, acompañadas de información acerca del tema pertinente y del grado de dificultad esperado. La información que se encuentra en un banco de preguntas se puede utilizar para crear pruebas de examen. En la actualidad, el IB no utiliza bancos de preguntas.
Bandas de puntuación	Rango específico de puntos que se deben otorgar a la respuesta de un alumno/a que demuestre ciertas cualidades.
Calificación	Descripción del logro de un alumno/a. Las calificaciones finales del trabajo entregado para evaluación van del 1 (la más baja) al 7 (la más alta). La calificación representa la valoración del IB de las cualidades generales que ha demostrado el alumno/a, y es coherente entre años y asignaturas. Una calificación describe la calidad del desempeño del alumno/a, y debe significar lo mismo en todos los exámenes, cursos y asignaturas. Esto tiene en cuenta la dificultad relativa del componente entre convocatorias de examen y cualquier cambio en los estándares.
Carpetas electrónicas	Sistema o proceso mediante el cual los colegios cargan el material de examen o de trabajo de clase evaluado internamente del alumnado del PAI para que el IB lo modere externamente.
Ciclo de evaluación	Pasos que se siguen para crear, efectuar y corregir las evaluaciones, incluidos los exámenes. Se trata de un ciclo porque el IB aprende de cada convocatoria de exámenes para mejorar las convocatorias futuras.
Circunstancias adversas	Circunstancias ajenas a la voluntad del alumno/a que puedan perjudicar su desempeño en la evaluación. Esto incluye situaciones de estrés grave,

Término	Definición
	<p>circunstancias familiares excepcionalmente difíciles, pérdida de un ser querido o sucesos que puedan poner en peligro la salud o la seguridad del alumnado. Las mismas circunstancias pueden afectar a un grupo de estudiantes, o todo el alumnado de un colegio. Entre las circunstancias adversas no se incluyen las siguientes:</p> <ul style="list-style-type: none"> • Deficiencias del colegio donde esté matriculado el alumno/a • La incapacidad del alumno/a para mejorar su desempeño a pesar de habersele autorizado adecuaciones inclusivas de acceso
Circunstancias excepcionales	<p>Circunstancias que comúnmente no se corresponden con las de otros alumnos/as con necesidades específicas de acceso a la evaluación. El IB se reserva el derecho de determinar qué circunstancias se pueden considerar “excepcionales” y, por lo tanto, justifican la aplicación de adecuaciones de acceso.</p>
Componente de evaluación	<p>Una o varias tareas que, conjuntamente, forman parte de la evaluación general, como un examen, una carpeta de trabajos, un proyecto, o un trabajo de investigación.</p>
Comprobador de compatibilidad	<p>Aplicación independiente que se utiliza en los exámenes en pantalla del PAI y en los exámenes digitales del PD y el POP para confirmar si un dispositivo cumple los requisitos técnicos para realizar un examen digital. El comprobador debe ejecutarse en todos los dispositivos que se utilizarán para la evaluación. No se pueden realizar exámenes en dispositivos que no superen la comprobación de compatibilidad.</p>
Comprobador de requisitos del sistema	<p>Herramienta que se utiliza para verificar que un dispositivo cumple los requisitos técnicos para realizar un examen digital del PD o el POP. Esto ayuda a los colegios a probar los dispositivos con antelación y a confirmar que permitirán una experiencia de evaluación fluida y segura.</p> <p>Este comprobador se utiliza durante la preparación de exámenes, no el día del examen, y complementa a los requisitos del sistema publicados.</p>
Concesión de calificaciones	<p>Proceso mediante el cual se determina cómo convertir las puntuaciones en calificaciones finales. Esto garantiza que las calificaciones finales tengan el mismo significado, independientemente de la convocatoria en la que el alumno/a realice el examen.</p>
Conducta impropia	<p>Toda acción (ya sea deliberada o involuntaria) de un alumno/a por la cual este o un tercero sale o puede salir beneficiado injustamente en uno o varios componentes de la evaluación. Las acciones que pudieran perjudicar a otros alumnos/as también se consideran conducta impropia.</p>
Consola de administración	<p>Sitio web seguro utilizado para los exámenes en pantalla del PAI. Permite al coordinador/a del PAI o al supervisor/a jefe designado realizar actividades administrativas clave, como el registro de dispositivos y el acceso a los paquetes de examen. El acceso a la consola de administración se realiza a través de un enlace en IBIS.</p> <p>Nota: Para los exámenes digitales del PD y el POP, véase sistema de exámenes digitales.</p>
Consulta sobre los resultados	<p>Revisión de la puntuación llevada a cabo a petición del colegio.</p>
Convocatoria	<p>Véase convocatoria de exámenes.</p>

Término	Definición
Convocatoria de exámenes	Período durante el cual se realizan y corrigen los exámenes. El IB celebra dos convocatorias de examen al año: una en mayo y otra en noviembre.
Corrección automática	Proceso de utilizar la tecnología para evaluar el trabajo del alumnado según un esquema de calificación predefinido en el que una respuesta es objetivamente correcta o incorrecta.
Corrección electrónica	Proceso mediante el cual los examinadores corrigen el material de examen directamente en la pantalla del dispositivo.
Correspondencia	Concordancia, en la teoría y la práctica, entre los valores y las aspiraciones que se comparten en lo relativo al aprendizaje (el currículo escrito), al modo de trabajo del profesorado (el currículo enseñado) y a lo que el alumnado aprende (el currículo evaluado).
Creador/a de contenidos de evaluación	Profesional que crea, verifica y revisa las preguntas de examen propuestas y el esquema de calificación asociado que se utilizará para las evaluaciones. Los creadores/as de contenidos de evaluación son casi siempre examinadores/as supervisores, si bien este trabajo requiere una serie de habilidades distintas.
Criterios de evaluación	Criterios en relación con los cuales se mide el desempeño de un alumno/a.
Criterios holísticos	Enfoque de evaluación que considera el trabajo del alumno/a como un resultado único, en lugar de analizar distintos elementos individuales de dicho trabajo (por ejemplo, la comunicación, el conocimiento de la materia, la calidad del argumento).
Cuestionario de examen	Conjunto de tareas y preguntas que se pide al alumnado que complete en un examen. Este término puede referirse a un examen que se responde en la pantalla de un dispositivo informático o a un examen que se responde en papel.
Curso	Número prescrito de clases, lecciones u horas lectivas en un período de estudio definido. Los colegios organizan el aprendizaje y la enseñanza de las asignaturas por medio de cursos disciplinarios e interdisciplinarios.
Descriptor de calificaciones finales	Explicación de las cualidades que se esperan en el alumnado para alcanzar cada calificación. Un descriptor de calificaciones finales puede ser específico de una asignatura, específico de un grupo de asignaturas, o general de todo un programa. En cada caso, un descriptor de calificaciones finales debe describir las mismas características; los ejemplos más específicos solo explican qué significan dichas descripciones en el contexto concreto de una asignatura.
Descriptor de nivel de logro	Descripciones de las características del trabajo entregado para evaluación que se esperan en cada nivel de logro.
Diferenciación (en el aprendizaje y la enseñanza)	Modificación de las estrategias de enseñanza para atender las diversas necesidades del alumnado por medio de distintos contenidos, procesos y productos.
Diferenciación (en la evaluación)	Distinción entre estudiantes que demuestran diferentes niveles de competencia.
Disciplina	Rama del aprendizaje o campo de estudio académico; forma de ordenar el conocimiento a efectos de la instrucción (generalmente denominada <i>asignatura</i> en el PAI y el PD para fines prácticos de evaluación).
Diseño universal de evaluación	Concepto según el cual todas las evaluaciones se deben desarrollar teniendo en cuenta la variedad de necesidades que puede tener el alumnado, en vez de

Término	Definición
	tratar a parte del alumnado de manera distinta. Esto forma parte del compromiso del IB con el diseño universal para el aprendizaje.
Diseño universal para el aprendizaje	Orientación y marco para el aprendizaje que se centra en la totalidad del alumnado. Su objetivo es crear entornos inclusivos, para lo cual proporciona múltiples medios de participación, representación y expresión.
Dispositivo	Computadora portátil o de escritorio que se utiliza para las evaluaciones digitales del IB. El IB utiliza el término <i>dispositivo</i> en lugar de <i>computadora</i> para reflejar la exploración continua de la compatibilidad futura con tabletas y otras plataformas.
Efecto de repercusión	Impacto que partes posteriores de un proceso tienen sobre partes anteriores. En el contexto educativo, este término normalmente se refiere al modo en que la evaluación afecta al aprendizaje y a la enseñanza.
Elemento	La subdivisión más pequeña apropiada de una evaluación; una parte de la evaluación independiente, pero completa, que se examina tal y como se describe en las secciones “Resumen del programa de estudios” o “Descripción detallada de la evaluación externa” de la correspondiente guía de la asignatura del IB.
Enfoque basado en criterios (véase también enfoque basado en logros)	Comparación del desempeño del alumno/a con descripciones predefinidas de logro (criterios) para realizar la corrección.
Enfoque basado en logros (también enfoque flexible basado en criterios)	Comparación del desempeño del alumno/a con descripciones predefinidas de logro (criterios) y el desempeño de anteriores grupos de estudiantes. Este es el enfoque que el IB adopta para mantener los estándares.
Enfoque flexible basado en criterios (también enfoque basado en logros)	Comparación del desempeño del alumno/a con descripciones predefinidas de logro (criterios) y el desempeño de anteriores grupos de estudiantes. Este es el enfoque que el IB adopta para mantener los estándares.
Enfoque normativo	Práctica por la cual el logro se determina al comparar el desempeño del alumno/a con el del grupo de estudiantes para quienes se diseñó la evaluación.
Entrega	Proceso por el cual un alumno/a (o el colegio en su representación) envía un trabajo final al IB.
Equiparabilidad	Grado en que un resultado concreto puede considerarse idéntico a otro. Se aplica generalmente entre años o entre asignaturas.
Esquema de calificación	Orientación para la concesión de los niveles de los criterios a un trabajo concreto.
Esquema de calificación analítico	Esquema de calificación que indica cuál es la respuesta correcta y en qué casos se deben conceder puntos.
Estandarización	Proceso colaborativo en el que los moderadores/as o examinadores/as acuerdan un estándar de evaluación común.
Estandarización interna	Proceso mediante el cual todo el profesorado de un colegio que imparte una asignatura concreta se asegura de que está aplicando el mismo estándar en la evaluación.

Término	Definición
Estrategia de evaluación	Método o enfoque empleado para obtener información sobre el aprendizaje del alumnado (por ejemplo, observación, tareas abiertas o respuestas seleccionadas).
Estudiante (también alumno)	Persona que estudia un curso o un programa del IB.
Estándar	Nivel de desempeño esperado que corresponde a una puntuación, una calificación o un resultado concreto.
Evaluación	Cualquiera de las tareas que completa un alumno/a para demostrar sus capacidades. Estas tareas pueden ser creación del colegio o del IB, y constituyen un conjunto de pruebas que permiten realizar valoraciones sobre el aprendizaje y la enseñanza.
Evaluación basada en criterios	Sistema de evaluación donde los niveles de logro se determinan en función de criterios previamente establecidos. Así pues, el estándar no es variable ni depende del nivel de logro alcanzado por todo el grupo de estudiantes, sino que es fijo.
Evaluación electrónica	Evaluación del PAI que consta de tres componentes: la evaluación en pantalla, la carpeta electrónica, y el Proyecto Personal.
Evaluación externa	Evaluación establecida y corregida por el IB, no por el profesorado del alumnado.
Evaluación final	Evaluación sumativa del trabajo al término del período de estudio.
Evaluación formativa	Evaluación continua del trabajo del alumnado que sirve para orientar la enseñanza y mejorar el desempeño de este.
Evaluación interdisciplinaria	Evaluación que combina o relaciona dos o más ramas del aprendizaje o campos de estudio académico. En el PD, las asignaturas interdisciplinarias son aquellas que reúnen los requisitos de dos grupos de asignaturas en una sola asignatura. En el PAI, el estudio interdisciplinario puede desarrollarse tanto dentro de los grupos de asignaturas como entre ellos. La evaluación interdisciplinaria externa del PAI siempre incluye varios grupos de asignaturas.
Evaluación interna	Evaluación llevada a cabo por el profesorado en el colegio.
Evaluación sumativa	Evaluación destinada a determinar la competencia o el nivel de logro de un alumno/a, y que normalmente tiene lugar al completar un programa de estudios o una unidad de trabajo.
Examen	Un examen es una forma de evaluación, y abarca una o varias tareas de diversos tipos (de respuesta breve, de respuesta amplia, de resolución de problemas o analíticas; a veces prácticas u orales) que el alumnado debe realizar en condiciones estrictamente controladas y con un límite de tiempo.
Examen con las respuestas del alumno o tarea de evaluación enviada	Respuestas del alumno/a a un examen en papel. También puede tratarse de un trabajo entregado para evaluación.
Examen en pantalla del PAI	Evaluación digital multimedia formal y con límite de tiempo diseñada específicamente para el PAI. Estos exámenes se llevan a cabo en un entorno digital seguro, y consisten en tareas específicas de cada asignatura que cada estudiante completa en un dispositivo. No es necesario disponer de conexión a Internet para realizar el examen, pero se recomienda encarecidamente hacerlo en línea. Esto permite que las

Término	Definición
	respuestas de cada estudiante se guarden automáticamente y reduce la carga de trabajo administrativo. Por este motivo, la mayoría de los colegios optan por realizar estas evaluaciones en línea.
Examinador	Persona que concede puntuaciones a las tareas de evaluación externa del alumnado.
Examinador jefe	El examinador/a supervisor con más experiencia que es responsable de garantizar el mantenimiento de los estándares a lo largo del tiempo y entre las disciplinas de un mismo grupo de asignaturas.
Examinador principal	<p>Persona responsable de dirigir la evaluación de un componente. El examinador/a principal establece el estándar de corrección y normalmente es también uno de los autores/as de la evaluación.</p> <p>En el PAI, la función del examinador/a principal difiere ligeramente con respecto a otros sistemas de evaluación. En el PAI, el examinador/a principal es responsable de una disciplina concreta y se encarga de dirigir al equipo que diseña la evaluación, definir y mantener los estándares y asesorar a los jefes/as de los equipos examinadores.</p>
Examinador supervisor	Examinador/a con experiencia que ayuda al examinador/a principal.
Exámenes de muestra	<p>Ejemplos completos de exámenes de asignaturas del IB. Los exámenes de muestra albergan contenido real de evaluación y reflejan el formato y el nivel de las evaluaciones finales. Normalmente se utilizan para ayudar al alumnado a practicar en condiciones similares a las del examen.</p> <p>PAI: En el Centro de Recursos para los Programas hay disponibles paquetes de muestra.</p> <p>PD y POP: En el Centro de Recursos para los Programas se publicarán exámenes de muestra. En el caso de los exámenes digitales, se puede proporcionar acceso adicional a través del sistema de exámenes digitales.</p>
Exámenes digitales	<p>En el PAI, los exámenes en pantalla se llevan realizando desde hace más de una década. Están diseñados específicamente para el entorno digital, e incluyen tareas interactivas y con contenidos multimedia que se completan en un entorno de examen seguro.</p> <p>En el PD y el POP, los exámenes digitales son evaluaciones formales que el alumnado realiza el último año y utilizando un dispositivo en condiciones de examen seguras. Se introducirán gradualmente, y supondrán una réplica digital de los exámenes tradicionales en papel.</p>
Factor de moderación	Ajuste aritmético aplicado a las puntuaciones concedidas por el profesor/a (en un nivel de un criterio) para adecuarla al estándar definitivo.
Familiarización	<p>Proceso mediante el cual el alumnado aprende a utilizar el entorno y las herramientas de examen digitales del IB, para lo cual se emplea una versión de práctica de los sistemas que no tiene contenido. Los colegios son responsables de garantizar que el alumnado tenga acceso regular a las herramientas de familiarización antes de sus exámenes digitales.</p> <p>En el Centro de Recursos para los Programas hay disponibles entornos de familiarización tanto para el PAI como para el PD y el POP.</p>
Fiabilidad	Medida en que el alumno/a logrará el mismo resultado en cada evaluación de su trabajo. Puede hacer referencia a la fiabilidad entre examinadores/as o a la fiabilidad de un único examinador/a.

Término	Definición
Grupo de preguntas	Se considera un grupo a una o varias preguntas de un cuestionario de examen que estén relacionadas entre sí. Los examinadores/as deben corregir los grupos de preguntas individuales, en lugar de exámenes completos. Este enfoque ofrece unos resultados más fiables que la corrección de exámenes completos.
Herramienta de comprobación del ancho de banda	Recurso en línea que utiliza el personal de dirección de un colegio del PD o del POP para confirmar que el ancho de banda de Internet del colegio es suficiente para que varios alumnos/as puedan realizar los exámenes digitales del PD o del POP al mismo tiempo. Se recomienda que el ancho de banda se compruebe con suficiente antelación a las convocatorias de exámenes para garantizar una configuración óptima.
Herramienta de evaluación	Método para obtener información sobre el desempeño y la comprensión de un alumno/a.
Herramienta de familiarización	Simulación sin contenido del entorno de examen digital. Esto permite al alumnado practicar cómo utilizar la plataforma —explorar las herramientas y comprender el formato del examen— sin ningún contenido específico de la asignatura.
Integridad académica	Conjunto de valores y habilidades que promueven la integridad personal y las buenas prácticas en el aprendizaje, la enseñanza y la evaluación.
Jefe de equipo	Examinador/a que dirige un equipo examinador.
Lengua de respuesta	Lengua que un alumno/a emplea para responder las preguntas en la evaluación.
Lenguas de trabajo	Lenguas en las que el IB se comunica con la comunidad escolar y en las que se compromete a ofrecer una serie de servicios para la implementación de los programas. Actualmente, las lenguas de trabajo del IB son español, francés e inglés.
Límite de calificación	El punto en el que el nivel de logro de un alumno/a pasa de una calificación a otra. A menudo se utiliza para indicar la puntuación total mínima y máxima que corresponde a una calificación concreta.
Mala administración	Acción por parte de un Colegio del Mundo del IB que infringe las normas del IB y que puede poner en peligro la integridad de los exámenes y de las evaluaciones del IB. Puede suceder antes, durante o después de completarse la evaluación o el examen.
Manejabilidad	Grado en el que una evaluación y cada una de las tareas suponen una carga para el alumno/a o el colegio. Algunos ejemplos de factores que se deben tener en cuenta para determinar la manejabilidad son la duración de las evaluaciones, el equipo o el material necesario para entregar la evaluación, o el número de evaluaciones que se requieren en un programa.
Margen de tolerancia	Pequeña desviación de la puntuación definitiva concedida por el examinador/a principal que el IB considera lo suficientemente próxima como para demostrar que el examinador/a está aplicando el estándar de corrección correcto. Los márgenes de tolerancia son necesarios, ya que la corrección no deja de ser una cuestión de valoración: incluso los examinadores/as con más experiencia pueden volver a corregir el mismo trabajo de una forma ligeramente distinta. El margen de tolerancia varía en función del número de puntos, el tipo de pregunta y la asignatura.
Material de ayuda al profesor	Información adicional para ayudar al profesorado a entender los requisitos del curso del IB. La finalidad del material de ayuda al profesor es proporcionar

Término	Definición
	orientación práctica y ayudar a comprender e implementar la teoría que se presenta en las guías de las asignaturas.
Matriculación de estudiantes	Proceso que lleva a cabo el coordinador/a de un programa para inscribir al alumnado para la evaluación del IB.
Modelo de calidad	Enfoque adoptado por el IB para garantizar que el alumnado recibe el resultado correcto de la evaluación. El examinador/a principal define el estándar de respuesta correcto para cada pregunta y todos los examinadores/as deben reproducirlo. En el caso de los trabajos evaluados externamente, el proceso implica proporcionar orientación a los examinadores/as por medio de la estandarización, comprobar su comprensión del estándar mediante respuestas de aptitud y supervisar regularmente su corrección por medio de respuestas de control.
Moderación	Proceso para garantizar la aplicación de un estándar de evaluación común. Este proceso implica la revisión de muestras de trabajos entregados para evaluación, y el ajuste de las puntuaciones otorgadas, según sea necesario.
Moderación externa	Véase moderación .
Modo restringido	Entorno de examen seguro que impide al alumnado acceder a otras aplicaciones, sitios web o archivos durante un examen digital. PAI: La función de modo restringido está integrada en el paquete de examen y se activa automáticamente cuando se inicia la evaluación. PD y POP: El modo restringido se aplicará mediante el navegador Safe Exam Browser para Windows y Mac, y el modo Kiosco de Chrome para las computadoras Chromebook gestionadas por el colegio. Los colegios son responsables de configurar los dispositivos del alumnado.
Muestra para la moderación	Muestra de trabajos que se entrega al IB para asegurarse de que está corregida de acuerdo con el estándar requerido.
Nivel de logro	Valor numérico que se otorga al trabajo entregado para evaluación cuando este refleja el nivel descrito por el descriptor al cual dicho valor corresponde. Se indica en la columna ubicada a la izquierda en los criterios de evaluación.
Objetivo específico	Cada uno de los enunciados que describen las habilidades, los conocimientos y la comprensión que se van a evaluar.
Pertinencia del constructo	Grado en el que la evaluación permite comprobar realmente los conocimientos y habilidades en cuestión.
Plagio	Presentación (ya sea intencional o no) de las ideas, las palabras o el trabajo de otra persona sin mencionarla de forma adecuada, clara y explícita.
Práctica indebida	Cualquier práctica contraria a los principios de integridad académica (por ejemplo, el plagio).
Predictibilidad	Grado en que es posible estimar qué sucederá o cuándo sucederá algo. En el ámbito de la evaluación, <i>predictibilidad</i> se refiere a la capacidad de los colegios para prever qué preguntas aparecerán en los exámenes y cuándo.
Pregunta	Tarea o actividad utilizada para que el alumnado demuestre su competencia en una asignatura.
Pregunta de opción múltiple	Pregunta en la que el alumno/a debe seleccionar la respuesta correcta de una lista de posibles respuestas que se proporcionan.

Término	Definición
Procedimiento de ausencia de calificación	Mecanismo para otorgar una calificación al alumnado del PAI en los casos en los que el IB no puede calcular una calificación justa o precisa basada en el trabajo que ha completado el alumno/a. Este proceso está pensado para aquellas circunstancias en las que la ausencia de pruebas se debe a alguna acción del IB o de terceras partes (a excepción del colegio), a consecuencia de la cual no sería razonable exigir al alumno/a que completara la evaluación en otra ocasión.
Procedimiento de ausencia de puntuación	Mecanismo para otorgar una puntuación al alumnado del PD y el POP en los casos en los que el IB no puede calcular una puntuación justa o precisa basada en el trabajo que ha completado el alumno/a. Este proceso está pensado para aquellas circunstancias en las que la ausencia de pruebas se debe a alguna acción del IB o de terceras partes (a excepción del colegio), a consecuencia de la cual no sería razonable exigir al alumno/a que completara la evaluación en otra ocasión.
Programa “Traiga su propio dispositivo”	Práctica en la que el alumnado utiliza sus propios dispositivos para su trabajo escolar. Los colegios proporcionan especificaciones para orientar a padres, madres y tutores/as legales sobre la compra de dispositivos adecuados para que el alumnado los utilice en el colegio.
Prueba	Véase cuestionario de examen .
Prueba	La palabra <i>prueba</i> se puede utilizar genéricamente como sinónimo de <i>examen</i> , a pesar de que a veces se le da un significado específico en trabajos académicos sobre el tema de la evaluación. En ocasiones, la evaluación general de un alumno/a se divide en partes que se realizan en momentos distintos. Estas se denominan <i>componentes de evaluación</i> o <i>componentes</i> . Algunos ejemplos de componentes pueden ser los exámenes orales, cualquier examen individual, o una evaluación interna.
Prueba modificada	Cambios efectuados en una evaluación para permitir que un alumno/a con necesidades específicas pueda realizarla en las mismas condiciones que el alumnado que no presenta tales necesidades. La modificación de la forma o el estilo de la fuente puede ser un ejemplo de estos cambios. Los ajustes realizados no deben alterar la naturaleza de la pregunta formulada.
Publicación de resultados	Proceso mediante el cual el alumnado recibe las calificaciones del IB correspondientes a sus evaluaciones.
Puntuación	Valor numérico que refleja la calidad de la respuesta de un alumno/a a una tarea específica. Las puntuaciones se otorgan de acuerdo con criterios, bandas de puntuación o un esquema de calificación para reflejar en qué medida el alumno/a ha respondido correctamente a una pregunta. La asignación de puntos varía en función de las preguntas y de los exámenes.
Puntuación definitiva	Puntuación que el examinador/a principal concede a un trabajo concreto entregado para evaluación. Representa la puntuación que todos los demás examinadores/as deberían intentar replicar. (Véase también modelo de calidad).
Repetidor	Categoría de matriculación para estudiantes que se presentan nuevamente a uno o varios exámenes con el objetivo de obtener el certificado del PAI o el diploma del IB, o de mejorar la puntuación total del certificado del PAI o del diploma del IB que ya han obtenido.
Requisitos del sistema	Especificaciones técnicas mínimas que debe cumplir un dispositivo para realizar los exámenes digitales del IB de forma fluida y segura. Estos requisitos

Término	Definición
	<p>se han establecido para garantizar que el alumnado no se vea perjudicado por problemas de hardware durante sus evaluaciones.</p> <p>Los requisitos se publican tanto para el PAI como para el PD y el POP. Los colegios son responsables de asegurarse de que todos los dispositivos del alumnado cumplan con las especificaciones necesarias.</p>
Respuesta atípica	<p>Respuesta a una tarea que difiere considerablemente de las respuestas habituales. Ejemplos de este tipo de respuestas pueden ser los trabajos incompletos o que no se han realizado conforme a las instrucciones, las respuestas imprevistas, los trabajos problemáticos o los casos de prácticas indebidas.</p>
Respuesta de aptitud	<p>Un ejemplo de trabajo entregado para evaluación que selecciona el examinador/a principal para comprobar que los examinadores/as han comprendido el estándar requerido de corrección antes de que corrijan en condiciones reales.</p>
Respuesta de control	<p>Examen corregido por el examinador/a principal y que se incluye aleatoriamente en un lote de exámenes asignado a un examinador/a para su corrección. Los puntos que el examinador/a concede a la respuesta de control se comparan con los concedidos por el examinador/a principal, para comprobar si el examinador/a está aplicando el estándar de corrección prefijado. Las respuestas de control de moderación de la toma dinámica de muestras se utilizan del mismo modo como parte del proceso de moderación.</p>
Respuesta de evaluación	<p>Término utilizado para describir todo el material elaborado por un alumno/a en respuesta a una tarea de evaluación.</p>
Respuesta de práctica	<p>Ejemplos de trabajos entregados para evaluación que se identifican y corrigen durante la estandarización. Estos se entregan al equipo examinador para que conozca el estándar que se debe aplicar en su corrección.</p>
Respuestas en los exámenes digitales	<p>En las evaluaciones digitales del IB, el trabajo que completa un alumno/a se denomina de forma diferente según el programa.</p> <p>En el PAI, se denomina <i>archivo de respuestas del alumno/a</i>.</p> <p>En el PD y el POP, se denomina <i>respuesta del alumno/a</i>.</p> <p>Ambos términos se refieren al producto digital final generado durante el examen y entregado para evaluación. En los exámenes en papel, esto se denomina <i>examen con las respuestas del alumno/a</i> o <i>examen escrito del alumno/a</i>.</p>
Resultados de los cursos	<p>Documento principal del IB en el que se informa los resultados de evaluación de cada estudiante. Muestra todas las asignaturas estudiadas por el alumno/a y la calificación final que ha logrado (1-7, o A-E cuando corresponda). También muestra otros resultados de evaluación pertinentes, tal como la calificación recibida en los componentes troncales.</p> <p>Por último, el documento incluye el nombre del alumno/a, su código personal, el número de convocatoria, la convocatoria en la que se han concedido las calificaciones, la fecha de expedición y el nombre del colegio que ha matriculado al alumno/a (además de un duplicado, si procede).</p>
Reunión de estandarización	<p>Reunión de los examinadores/as principales para describir el estándar de corrección que debe aplicarse y definir las respuestas de control.</p>

Término	Definición
Revisión de la corrección del examinador	Proceso de revisión de la corrección de las respuestas del alumnado asignadas a un examinador/a (ya sea respuestas completas o grupos de preguntas, dependiendo de la estructura de la evaluación) si se detecta que la corrección no se ajusta al estándar requerido o difiere considerablemente de este. Con frecuencia, esto se debe a un error en la moderación.
Sesgo	Efecto por el cual un grupo definido tiene, en una pregunta o tarea específica, un desempeño distinto al promedio por algún motivo que no sea su capacidad en la habilidad o el conocimiento que se evalúa.
Sistema de exámenes digitales	Sistema en línea que los colegios utilizan para administrar los exámenes digitales del PD y el POP.
Sistema de Información del IB (IBIS)	Sistema que permite a los coordinadores/as de los programas realizar procedimientos administrativos y obtener noticias e información del IB por medio de un servidor web al que se accede con una contraseña.
Supervisor de exámenes	Persona que supervisa el entorno de examen y ayuda a mantener la seguridad del examen.
Supervisor jefe	<p>PD y POP: El supervisor/a jefe es responsable de supervisar los exámenes digitales en el colegio, lo que incluye completar tareas administrativas en el sistema de exámenes digitales y gestionar cualquier problema técnico antes y durante el período de exámenes. Debe ser miembro del personal del colegio, y el coordinador/a del programa le asigna esta función en Mi IB.</p> <p>Esta función debe ser independiente de la del coordinador/a del programa, pero quien la desempeña debe trabajar en estrecha colaboración con este para garantizar que los exámenes se desarrollen según lo previsto. En algunos casos, el coordinador/a del programa también puede desempeñar la función de supervisor/a jefe.</p> <p>PAI: La función de supervisor/a jefe es opcional. El coordinador/a del programa puede optar por asignar a alguien para que ayude a completar las tareas administrativas relacionadas con los exámenes en pantalla.</p>
Tarea de evaluación	Actividad o serie de actividades que realiza el alumnado a fin de que tenga lugar la evaluación.
Término de instrucción	Palabra o palabras de una pregunta que explican el objetivo de evaluación que se evalúa.
Toma dinámica de muestras	Perfeccionamiento del proceso de moderación que permite que se aprovechen mejor los controles de calidad.
Totalización	Proceso de combinar las puntuaciones obtenidas para obtener un resultado final.
Trabajo evaluado externamente	Trabajo evaluado íntegramente por el IB.
Trabajo evaluado internamente	Trabajo evaluado por los profesores/as del alumnado. Con el fin de realizar la moderación, el IB pide muestras del trabajo evaluado internamente.
Validez	Término que describe si una evaluación, o si el propósito para el que se van a usar los resultados de la evaluación, son adecuados.
Valoración	Consideración del trabajo de un alumno/a en relación con un criterio de evaluación determinado.

Término	Definición
Verificación de la autoría original	Pruebas y proceso de comprobación de que el trabajo lo ha llevado a cabo el propio alumno/a. Algunos ejemplos pueden ser las firmas del profesor/a y del alumno/a que confirman que el trabajo es suyo.

Recursos para imprimir

A lo largo de esta publicación se hacen referencias a varios recursos para imprimir. A continuación se enumeran todos los recursos para imprimir, con el fin de facilitar el acceso a ellos.

- [“Calificaciones previstas: guía para el profesorado” \(PDF\)](#)
- [“Corrección de la evaluación interna: expectativas del profesorado en el proceso de moderación” \(PDF\)](#)
- [“Desarrollo de la mentalidad ética” \(PDF\)](#)
- [“Evaluación sumativa y evaluación formativa” \(PDF\)](#)
- [“Implementación de adecuaciones inclusivas de acceso” \(PDF\)](#)
- [“La cadena de validez” \(PDF\)](#)
- [“Los principios de evaluación del IB” \(PDF\)](#)
- [“Moderación con toma dinámica de muestras ” \(PDF\)](#)
- [“Rango de trabajos de evaluación interna necesario para el modelo de calidad” \(PDF\)](#)
- [“Responsabilidades de quienes desempeñan funciones clave en el ciclo de evaluación” \(PDF\)](#)
- [“Taxonomía de Bloom” \(PDF\)](#)