

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





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

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

Versión en español del documento publicado en julio de 2019 con el título
Assessment principles and practices—Quality assessments in a digital age
Publicada en julio de 2019
Actualizada en noviembre de 2021

Publicada en nombre de la Organización del Bachillerato Internacional, una fundación educativa sin fines de lucro con sede en 15 Route des Morillons, 1218 Le Grand-Saconnex, Ginebra (Suiza),

por

International Baccalaureate Organization Ltd (Reino Unido)
Peterson House, Malthouse Avenue, Cardiff Gate
Cardiff, Wales CF23 8GL
Reino Unido
Sitio web: ibo.org/es

© Organización del Bachillerato Internacional, 2019

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 los titulares de los derechos y obtener de ellos 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 uso del género masculino en esta publicación no tiene un propósito discriminatorio y se justifica únicamente como medio para hacer el texto más fluido. Se pretende que el español utilizado sea comprensible para todos los hablantes de esta lengua y no refleje una variante particular o regional.

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 terceros 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, proveedores de desarrollo profesional, empresas editoriales del sector educativo y compañías que ofrecen servicios de planificación curricular o plataformas digitales que brindan recursos a los docentes). 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 jóvenes solidarios, informados y ávidos 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 ser compasivos 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 ser:

#### **INDAGADORES**

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

#### **INFORMADOS E INSTRUIDOS**

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.

#### **PENSADORES**

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.

#### **BUENOS COMUNICADORES**

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

#### ÍNTEGROS

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.

#### **DE MENTALIDAD ABIERTA**

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

#### **SOLIDARIOS**

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

#### **AUDACES**

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.

#### **EQUILIBRADOS**

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

#### REFLEXIVOS

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. Estamos convencidos 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

introduccion y descripcion general	1
Introducción	1
¿Qué programas del IB cubre este recurso?	4
¿Por qué debe leer la publicación Principios y prácticas de evaluación?	5
¿Para quién está pensado este recurso?	6
Cómo utilizar este recurso: distintas maneras de sacarle provecho	7
¿Qué significa "evaluaciones de calidad en la era digital"?	12
¿Cómo está relacionado este recurso con otras publicaciones del IB?	13
La evaluación mediante el uso de la tecnología	16
¿Qué son los principios y prácticas de evaluación?	20
Terminología de la evaluación	21
Nuevos términos relacionados con la evaluación electrónica	23
Sección A: Principios de evaluación	25
¿En qué consiste la evaluación?	25
En qué consiste ser adecuado al propósito: el concepto de validez	31
Elementos de la cadena de validez	37
Definición de los estándares	54
La descripción del éxito: logros del alumno en la evaluación sumativa	59
Corrección de las evaluaciones	63
¿Qué es una buena evaluación?	69
Principios de evaluación del IB	81
Sección B: Las prácticas de evaluación del IB	82
¿Qué se entiende por <i>práctica</i> ?	82
Comunicación de los logros de los alumnos	83
El proceso de evaluación: funciones y responsabilidades	92
La integridad de la evaluación	97
Equidad para todos: cómo satisfacer las necesidades de los alumnos	101
El ciclo de evaluación	106
Preparación de pruebas de examen: desarrollo y calidad	109
Exámenes	114
Fijación del estándar: cómo preparar a los examinadores para la corrección	118
Corrección	123
Moderación	137
Concesión de calificaciones finales (y totalización)	144
Controles de calidad de casos en riesgo	158
El Comité de la evaluación final	159

Preparación para la publicación de los resultados	161	
Equidad para todos: cómo satisfacer las necesidades de los alumnos	166	
Consultas sobre los resultados, apelaciones y comentarios generales	171	
Preparación de las evaluaciones del año siguiente	176	
Comentarios para los colegios	179	
Sección C: Procesos específicos de los programas del IB	181	
¿En qué consisten los procesos específicos de los programas?	181	
Elementos comunes a todos los programas	183	
Las competencias del alumno y el perfil de la comunidad de aprendizaje del IB	185	
El Programa del Diploma del IB	191	
El Programa de Orientación Profesional del IB	198	
El Programa de los Años Intermedios del IB	202	
El Programa de la Escuela Primaria del IB	207	
Anexos	209	
Anexo 1: Moderación de la evaluación interna	209	
Anexo 2: Hoja de ruta para crear un argumento de validez	215	
Bibliografía	218	
Glosario		

#### Introducción

El objetivo de documento es explicar los principios que se emplean para garantizar que la evaluación que realiza el Bachillerato Internacional (IB) sea significativa, justa y beneficiosa para los alumnos.

Este recurso se redacta teniendo en mente a los profesores, pero también con la intención de que resulte accesible para los alumnos, los examinadores y otras partes interesadas.

Con la evaluación electrónica se puede evaluar lo que resulta importante, no solo aquello que es posible evaluar con papel y bolígrafo; además, utiliza el medio que mejor conoce la generación actual de alumnos.

No todo lo que se puede contar cuenta, y no todo lo que cuenta se puede contar.

Albert Einstein/William Cameron, 1963.

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.

1

Estas dos citas reflejan el alcance del desafío que plantea la evaluación. Muchos de los objetivos de la educación del IB no son fáciles de evaluar, pero si no se evalúa con precisión a los alumnos, entonces algunas decisiones importantes que afectarán a sus vidas se basarán en razones menos justas y menos fiables.

Los programas del IB se imparten en colegios de más de 140 países que representan una amplia variedad de tradiciones y contextos educativos. En algunos de estos contextos, la filosofía y los enfoques que adopta el IB en la evaluación de sus alumnos resultarán familiares, si bien para otros el sistema puede parecer misterioso y algo extraño.

Tener las cosas claras resulta aún más importante durante un período de cambios, y el impacto de la tecnología en la educación, incluida la evaluación, seguirá siendo notable en la próxima década. El IB cree firmemente que la tecnología debe estar al servicio de la evaluación. El avance hacia un sistema informatizado de evaluación en pantalla no cambiará los principios del IB, y además puede abrir nuevas posibilidades para convertir dichos principios en prácticas. Para obtener más información sobre este tema, consulte la sección "La evaluación mediante el uso de la tecnología".

Es importante que todos los componentes de la comunidad del IB comprendan cómo funciona nuestro proceso de evaluación externa, cuáles son sus puntos fuertes y sus limitaciones, y las razones por las que se toman las decisiones. Una mayor transparencia solo puede llevar a una mayor comprensión y, en última instancia, a una mejor educación para los alumnos. Gracias a las oportunidades que ofrecen los recursos informáticos, se espera proporcionar a los profesores una guía clara y accesible que contiene la información que necesitan para comprender la evaluación del IB.



Figura 1
Todas las evaluaciones requieren hallar un equilibrio

Todas las evaluaciones requieren hallar un equilibrio entre demandas opuestas, pero muchas veces los procesos de examen no tienen esto en cuenta.

Un ejemplo de esto es la tensión que surge entre el deseo de reducir la carga de las evaluaciones y el riesgo de que los alumnos solamente tengan una oportunidad para demostrar lo que saben hacer.

Para obtener más información sobre este tema, consulte la sección "En qué consiste ser adecuado al propósito: el concepto de validez".

# Necesidad de encontrar un equilibrio entre prioridades opuestas

La noción de *equilibrio* engloba otro aspecto: para el IB, el desarrollo de los alumnos pasa por un programa de estudio holístico, lo cual debe reflejarse en el enfoque de la evaluación. Esto significa que se deben tomar decisiones sobre cuál será el impacto sobre el programa en conjunto y no centrarse en una sola asignatura, disciplina o evaluación.

Para poder comprender la naturaleza de los principios de evaluación del IB y su funcionamiento, es necesario proporcionar algo de contexto sobre el desarrollo histórico y teórico de las prácticas de evaluación. Muchos temas importantes solo se tratan brevemente, pero es importante destacarlos, ya que tienen una gran influencia en las prácticas actuales. Para quienes quieran seguir leyendo sobre el tema, los trabajos académicos citados en el texto serán un buen punto de partida para seguir investigando.

Ya se ha indicado la dificultad de la tarea que se plantea el IB: centrarse en lo que es importante evaluar, y no en lo que es sencillo evaluar. Tal vez Alec Peterson, el primer director general del IB, lo expresó de la manera más elocuente:

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.

Se espera que el resto de este recurso logre explicar cómo el IB considera que puede estar a la altura de este objetivo tan ambicioso y respaldar sus intenciones educativas amplias para ofrecer a sus alumnos una experiencia de primer nivel. Si tiene alguna pregunta que este recurso no logre responder, escriba al personal de evaluación del IB a la dirección de correo electrónico assessment@ibo.org.

## ¿Qué programas del IB cubre este recurso?

Las versiones anteriores de los documentos dedicados a los principios y prácticas de evaluación se han centrado en el Programa del Diploma. Sin embargo, dada la ampliación de la variedad de programas del IB, resulta adecuado explicar en detalle la filosofía de evaluación del IB que se aplica en todos los programas: el Programa del Diploma (PD), el Programa de Orientación Profesional (POP), el Programa de los Años Intermedios (PAI) y el Programa de la Escuela Primaria (PEP). En este documento se tratarán en profundidad las razones en las que se basan los principios de evaluación del IB, pero no se abordará en detalle cómo dichos principios se aplican en cada programa. Si desea obtener este tipo de información, puede consultar el recurso De los principios a la práctica correspondiente al PAI, el PD o el POP.

Este documento se centra particularmente en la evaluación sumativa externa del IB. Actualmente, el IB ofrece una evaluación sumativa externa en el PD, el POP y el PAI.

Si bien la evaluación es un elemento importante del PEP, no resulta adecuado que cuente con una evaluación sumativa externa del IB. Para obtener más información sobre la filosofía de la evaluación en el PEP, consulte la sección de evaluación del PEP disponible en el Centro de recursos para los programas.

Los principios de evaluación del IB cubren una amplia variedad de propósitos para la evaluación, incluida la evaluación formativa, que conducen a las prácticas para los componentes que evalúa el IB. Estos principios resultarán de interés para todos los programas.



## ¿Por qué debe leer la publicación Principios y prácticas de evaluación?

## Porque los resultados de las evaluaciones influyen en la vida de los alumnos

La mayoría de los contenidos de esta publicación se refieren al modo en que el IB evalúa a los alumnos para conceder los resultados del Programa del Diploma (PD), del Programa de Orientación Profesional (POP) y, de manera opcional, del Programa de los Años Intermedios (PAI). Los alumnos, a su vez, utilizan estos resultados para progresar en su educación o comenzar su carrera profesional.

La evaluación es solo una herramienta que puede conllevar resultados positivos o negativos a quienes se les aplica. Como profesor, padre, madre o alumno, usted tiene un papel en las evaluaciones y por eso debe comprender sus puntos fuertes y puntos débiles, así como las decisiones que deben tomar las personas que ofrecen y utilizan las evaluaciones. A continuación encontrará algunas de las preguntas y comentarios frecuentes que recibimos por parte de profesores, y también la manera en que el presente documento les ayudará a resolver dichas cuestiones.

#### "¿Y por qué hace eso el IB?".

Con frecuencia oímos esta pregunta tan sincera por parte de profesores y alumnos cuando se enfrentan a la evaluación externa. La lectura de este documento debería servir para comprender los principios que rigen las prácticas de evaluación del IB.

#### "Pero esto no es justo, ¡necesito una calificación más alta!".

Los alumnos tienen mucho en juego en función de los resultados de sus exámenes, como su acceso a la universidad y su futura carrera profesional, y por eso es importante que comprendan por qué se toman las decisiones. La evaluación consiste en hallar un equilibrio entre demandas opuestas. Por eso, el presente documento le ayudará a explicar a los alumnos las implicaciones que tendría eludir las normas en sus casos, y tal vez sirva incluso para convencerles de que la calificación que recibieron fue justa.

#### "Nunca antes me lo había planteado así".

Es posible que incluso para algunos profesores que llevan años trabajando en el sector de la educación, la evaluación externa parezca un proceso misterioso y poco claro. Ninguna parte de la educación, especialmente si se trata de la educación del IB, se debería percibir como algo que "se le hace" a alguien, así que debemos ser capaces de explicar a los alumnos cómo y por qué reciben los resultados que reciben.

#### "Solamente quiero utilizar las evaluaciones para poder enseñar mejor".

Aunque la evaluación se quiera utilizar únicamente con fines formativos, es importante comprender en qué consiste una buena evaluación y qué decisiones hay que tomar. Las evaluaciones de poca calidad conducirán a resultados de poca calidad, y lo mismo sucede si el resultado es apoyar el aprendizaje o algún otro propósito escolar interno como, por ejemplo, una selección o un reconocimiento formal. Este documento le ayudará a entender cómo el IB cumple sus objetivos para ofrecer una evaluación de alta calidad a fin de respaldar sus metas educativas.

## ¿Para quién está pensado este recurso?

La idea de partida es que el lector es un profesor que está familiarizado con el IB y sus evaluaciones, pero que no ha estudiado formalmente la teoría de la evaluación. El IB considera que este postulado permite que este recurso sea accesible para el mayor público posible.

Es probable que este recurso resulte interesante a alumnos, padres, directores de colegios y coordinadores del IB; además, se proporciona un glosario y enlaces a otros documentos para lograr una comprensión plena de todos los procesos implicados.

De manera similar, las secciones que tratan sobre los principios en los que se basan las prácticas de evaluación del IB resultarán interesantes para comprender por qué se sigue el enfoque que se ha elegido. Las secciones de descripción general resultarán especialmente útiles para que los examinadores comprendan el papel que desempeñan en el contexto más amplio de la evaluación del IB.

Por último, gracias a este documento, los responsables de admisiones de las universidades y otras partes interesadas lograrán entender qué espera conseguir el IB mediante sus evaluaciones y qué representan los resultados de los alumnos.



## Cómo utilizar este recurso: distintas maneras de sacarle provecho

- En este documento, los temas están organizados de tal modo que los lectores se pueden dirigir a las secciones que les resulten de especial interés.
- Las preguntas sobre el tema que aparecen a continuación pueden servir como punto de partida para aprender sobre los aspectos que el lector considere más importantes.

Este documento de consulta se ha diseñado para utilizarse en función de los intereses y las necesidades del lector. El índice que aparece a continuación está pensado como lista de referencia de las secciones, pero no indica el orden recomendado para leerlas.

## ¿En qué consisten las preguntas sobre el tema?

El objetivo de este documento es aportar una visión general e integral del modo en que el IB aborda la evaluación. Muchos profesores no necesitarán tanta amplitud, sino información específica sobre determinadas partes del proceso. Para facilitar su búsqueda, se han preparado varias listas de secciones que se centran en preguntas concretas que los profesores podrían tener.

#### ¿En qué sentido la evaluación en pantalla supondrá una diferencia para la evaluación del IB?

Pensada para los profesores que están familiarizados con la evaluación del IB y que quieren saber cómo gestionará el IB el paso a las evaluaciones en pantalla.

Ejemplos de preguntas habituales:

- ¿En qué consiste la evaluación en pantalla?
- ¿La corrección la hará una computadora?
- ¿Por qué será mejor?
- ¿Se seguirán utilizando los mismos estándares?

La evaluación mediante el uso de la tecnología	¿Cómo es una evaluación en pantalla?	Riesgos de la evaluación en pantalla que se deben evitar
Terminología de la evaluación	Términos clave que es necesario comprender	Alumno matriculado y alumno
Evaluación, examen, prueba y componente	Puntuación y calificación	Pregunta y elemento
Autor del examen y examinador	Nuevos términos relacionados con la evaluación electrónica	Evaluación en pantalla
Archivo de respuestas y respuesta del alumno	Herramienta de familiarización	En caso de duda sobre algún término
¿Qué es una buena evaluación?	¿Cómo es una buena evaluación en pantalla?	El ciclo de evaluación

El impacto de la evaluación	El paso a la evaluación en pantalla	Evaluaciones en pantalla
electrónica en el ciclo de		
evaluación		

## ¿Cómo se corrigen las evaluaciones y cómo se otorgan las calificaciones?

Explicación para los profesores sobre qué sucede cuando envían los trabajos de los alumnos al IB para la corrección y sobre cómo se conceden las calificaciones.

Ejemplos de preguntas habituales:

- ¿Por qué este año los límites de calificación son distintos?
- ¿Por qué no puedo tener otro examinador?
- ¿Quién corrige los exámenes?
- ¿Cómo se comprueba el trabajo de los examinadores?
- ¿Cómo puedo presentar una apelación?
- Yo le hubiera concedido una puntuación de X al alumno; ¿por qué es esto justo?
- El esquema de calificación indica que la respuesta del alumno merece una puntuación de X; ¿cómo puede ser esto justo?
- ¿No sería mejor para los alumnos tener un "aula sin calificaciones"?

Terminología de la evaluación	Términos clave que es necesario comprender	Alumno matriculado y alumno
Evaluación, examen, prueba y componente	Puntuación y calificación Pregunta y elemento	
Comunicación de los logros de los alumnos	¿Qué significan las calificaciones finales del IB? ¿Cuál es la diferencia entre puntuaciones y calificacione	
La tiranía de las calificaciones: un mal menor	Qué se considera una convocatoria de exámenes satisfactoria	Los logros no se limitan a las calificaciones
El proceso de evaluación: funciones y responsabilidades	Examinador principal y examinador jefe	Otros examinadores
Las responsabilidades del personal del IB	Jerarquía de los examinadores	Elementos de la cadena de validez
Fijación del estándar: cómo preparar a los examinadores para la corrección	Reunión de estandarización	Exámenes de práctica
Exámenes de aptitud	Exámenes de control	Márgenes de tolerancia
Fijación satisfactoria del estándar	Corrección	Comprender la corrección como el resultado del juicio coherente de los examinadores
"Puntuaciones definitivas"	Puntuación y calificación son conceptos distintos	Principios básicos de la corrección en pantalla
Distintos tipos de esquemas de calificación	Esquemas de calificación analíticos	Criterios holísticos: bandas de puntuación



Apoyo adicional para los examinadores	Grupos de preguntas	Modelo de calidad
Exámenes de práctica	Exámenes de aptitud	Exámenes de control
Márgenes de tolerancia	Respuestas complicadas o inusuales	Vínculos con los colegios
Totalización	Concesión de calificaciones finales (y totalización)	En qué consiste la concesión de calificaciones finales
Límites de calificación estimativos e interpolados	El impacto de la evaluación electrónica en la concesión de calificaciones finales	Pruebas utilizadas en la concesión de calificaciones finales
Consideración del grupo de alumnos 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	Descriptores de calificaciones finales
Límites de calificación fijos	Totalización	Los controles de calidad en la concesión de calificaciones finales y los informes de distribución
Otorgamiento del título académico de un programa	Profesores observadores	Principios de la concesión de calificaciones finales
Controles de calidad de casos en riesgo		

#### ¿Cuál es el enfoque del IB con respecto a la evaluación?

Para el personal de admisiones de las universidades, profesores y partes implicadas que tienen interés en las bases teóricas de las evaluaciones del IB, la definición de evaluaciones de buena calidad, 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?

¿En qué consiste la evaluación?	¿Qué es la evaluación?	Evaluación formativa, sumativa y como aprendizaje: ¿por qué se evalúa?
El "efecto de repercusión" y el aprendizaje	Puntuación y calificación son conceptos distintos	En qué consiste ser adecuado al propósito: el concepto de validez
Significado de validez	¿Qué se considera válido: la evaluación o su uso?	Creación de un argumento de validez
Mantenimiento de la validez	Para la validez, ¿cuáles son las ventajas de trabajar en pantalla? ¿Y las ventajas de la evaluación electrónica?	Elementos de la cadena de validez
El equilibrio entre los distintos aspectos de la validez	Fiabilidad	Resultados coherentes no es lo mismo que resultados adecuados
Pertinencia del constructo y autenticidad	Manejabilidad	Equidad y sesgo

Equiparabilidad	El enfoque del IB acerca de la validez	Definición de los estándares
Los tres significados de estándares	El enfoque normativo y el enfoque basado en criterios en los estándares de desempeño	¿En qué consiste el enfoque normativo?
¿En qué consiste el enfoque basado en criterios?	¿Qué enfoque utiliza el IB?	Mantenimiento de los estándares
La descripción del éxito: logros del alumno en la evaluación sumativa	La tiranía de las calificaciones: un mal menor	La importancia del juicio profesional
Corrección de las evaluaciones	¿Qué se entiende por corrección?	Otras formas de corrección
Corrección y evaluación formativa	¿Qué es una buena evaluación?	Una buena evaluación sirve de apoyo a los objetivos curriculares
¿Qué es la 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 alumno y las habilidades de pensamiento de orden superior	Habilidades cognitivas de orden superior
Las competencias del alumno y el perfil de la comunidad de aprendizaje del IB	Mentalidad internacional y entendimiento intercultural	¿Cómo es una buena evaluación en pantalla?
Principios de evaluación del IB	Comunicación de los logros de los alumnos	¿Qué significan las calificaciones finales del IB?
¿Cuál es la diferencia entre puntuaciones y calificaciones?	La tiranía de las calificaciones: un mal menor	Qué se considera una convocatoria de exámenes satisfactoria
Los logros no se limitan a las calificaciones		

#### ¿Cómo diseña y redacta el IB los cuestionarios de examen?

Para profesores que quieren entender cómo se preparan los cuestionarios de examen e infundir confianza a aquellas personas a las que le preocupa la calidad de los cuestionarios de examen.

Ejemplos de preguntas habituales:

- ¿Quién redacta los cuestionarios de examen?
- ¿Cómo se comprueban?
- ¿Su contenido es el mismo cuando están en lenguas distintas?
- ¿Estos cuestionarios realmente comprueban lo adecuado?

¿En qué consiste la evaluación?	¿Qué es la evaluación?	Evaluación formativa, sumativa y como aprendizaje: ¿por qué se evalúa?
	Puntuación y calificación son conceptos distintos	En qué consiste ser adecuado al propósito: el concepto de validez



Significado de validez	¿Qué se considera válido: la evaluación o su uso?	¿Qué es una buena evaluación?
Una buena evaluación sirve de apoyo a los objetivos curriculares	¿Qué es la 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	Funciones relacionadas con la creación de las pruebas de examen	Preparación de pruebas de examen: desarrollo y calidad
Reglas que rigen la redacción de los cuestionarios de examen	Bancos de preguntas	Descripción general de la preparación de pruebas de examen
El proceso de aprobación de los recursos	El proceso de aprobación del contenido	El proceso de aprobación del formato
El proceso de aprobación de la usabilidad	Control de calidad	Traducciones
Cómo gestionar las solicitudes de pruebas de examen modificadas	El paso a la evaluación en pantalla	

#### ¿Cómo modera el IB las evaluaciones que corrigen los profesores?

Para profesores y partes interesadas que quieren entender por qué el IB incluye la evaluación de los profesores en sus procesos, pero posteriormente no siempre acepta las valoraciones que hacen los profesores.

Ejemplos de preguntas habituales:

- ¿Por qué han cambiado mis puntuaciones?
- ¿Por qué mi factor de moderación cambia de un año a otro?

¿En qué consiste la evaluación?	¿Qué es la evaluación?	Evaluación formativa, sumativa y como aprendizaje: ¿por qué se evalúa?
El "efecto de repercusión" y el aprendizaje	Puntuación y calificación son conceptos distintos	Moderación
En qué consiste la moderación	Selección de los trabajos de los alumnos	Trabajos inusuales o atípicos
Casos en los que no es posible encontrar un factor de moderación	Toma dinámica de muestras	El sistema anterior: la jerarquía de la moderación
Comentarios sobre la evaluación interna	Fijación del estándar: cómo preparar a los examinadores para la corrección	Reunión de estandarización
Exámenes de práctica	Exámenes de aptitud	Exámenes de control
Márgenes de tolerancia		

## ¿Qué significa "evaluaciones de calidad en la era digital"?

La tecnología es, cada vez más, una parte importante en nuestras vidas. Las formas de comunicación digital, como puede ser la mensajería a través de medios sociales, son hoy por hoy más comunes que escribir cartas, y en muchos empleos la computadora es una herramienta imprescindible. Asimismo, actualmente las computadoras y la tecnología en general tienen un impacto positivo en la enseñanza y el aprendizaje, y el IB está comprometido con ayudar a los colegios a que utilicen la tecnología de la mejor manera posible en sus aulas.

Hoy en día, muchos alumnos redactan sus ensayos y trabajos de investigación utilizando una computadora; de hecho, escribir en papel durante dos o tres horas se ha convertido en una tarea con la que están poco familiarizados. Por eso, los exámenes del IB deberían ser una oportunidad para que los alumnos demuestren qué entienden, en lugar de ser una experiencia excepcional que deben aprender a dominar.

El hecho incluir el lema "evaluaciones 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 los alumnos conozcan bien y en las que se sientan cómodos. 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 significa que la tecnología se debe adaptar a las necesidades de evaluación, y nunca al revés.

Comprender la evaluación en la era digital permite, sobre todo, mejorar aún más las evaluaciones. No se trata simplemente de evitar los efectos negativos al pasar de los exámenes y las evaluaciones en el aula (evaluaciones internas) a un formato electrónico, sino de utilizar la tecnología para superar las limitaciones que el IB lleva 50 años afrontando al confiar en los trabajos de clase y los exámenes en papel.

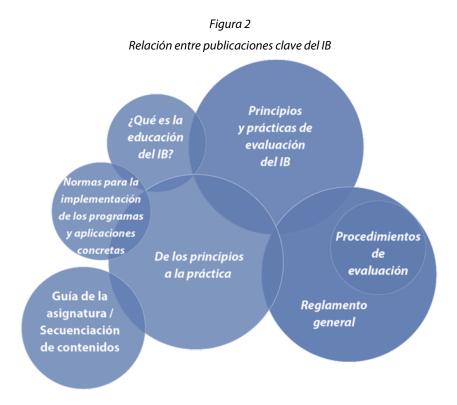
Poner la tecnología al servicio de la evaluación ofrece la oportunidad de aumentar la validez de las evaluaciones, y el lema subraya el entusiasmo del IB con este proceso, a la vez que refuerza su compromiso con los principios que sustentan todas sus evaluaciones.



## ¿Cómo está relacionado este recurso con otras publicaciones del IB?

Este recurso es uno de los muchos publicados por el IB para explicar su concepción de la educación. A veces puede resultar confuso entender cómo estos recursos clave están relacionados entre sí.

El siguiente diagrama ilustra dicha relación. La tabla, por su parte, identifica los distintos propósitos que tienen varios recursos clave del IB.



Publicaciones clave del IB que cubren todos los programas		
Principios y prácticas de evaluación del IB: evaluaciones de calidad en la era digital	Esta publicación explica el enfoque global que adopta el IB con respecto a la evaluación y cómo está pensado aplicarlo en la práctica. Se centra en el elemento de evaluación sumativa (evaluación formal) de la educación del IB.	
¿Qué es la educación del IB?	El propósito de este documento es informar sobre la esencia de la educación del IB y explicar los ideales que sirven de base a 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.	

Publicaciones clave del IB que cubren todos los programas	
Normas para la implementación de los programas y aplicaciones concretas	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. 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).

Publicaciones clave del IB específicas de los programas	
De los principios a la práctica	Esta publicación se centra en la enseñanza y el aprendizaje en el contexto de un programa concreto del IB. Además, explica los requisitos del programa.
Guías de las asignaturas (solamente PD y PAI)	Cada guía contiene información detallada sobre una asignatura, un curso o una disciplina; por ejemplo, Geografía o Artes Visuales.  En estos documentos 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 la enseñanza y el aprendizaje de la asignatura concreta.
Documentos de secuenciación de contenidos (solamente para el PEP)	El conjunto de documentos de secuenciación de contenidos del PEP ofrece ejemplos de cómo documentar las expectativas curriculares de cada área disciplinaria. Sus propósitos son: proporcionar una herramienta para informar a toda la comunidad escolar sobre la enseñanza y el aprendizaje en cada área disciplinaria, mostrar claramente los elementos esenciales del PEP en el contexto de las áreas disciplinarias, y aclarar el papel de las áreas disciplinarias en un programa transdisciplinario.
Procedimientos de evaluación	Define las reglas y procesos específicos que se deben seguir para la evaluación del programa del IB correspondiente.
Reglamento general	Conjunto de reglas que deben seguir los Colegios del Mundo del IB.
La realización de los exámenes del Programa del Diploma del IB (el POP utiliza la versión del PD) y La realización de los exámenes en pantalla del Programa de los Años Intermedios	Estas publicaciones informan a los coordinadores y a los supervisores de examen 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 estos documentos.

Además de estos valiosos recursos, el IB ofrece una variedad de recursos distintos que sirven de orientación para profesores, coordinadores y otras partes interesadas y que se basan en los principios que se establecen



en estas publicaciones clave. Estos recursos se encuentran en el sitio web del IB y en el Centro de recursos para los programas.

## La evaluación mediante el uso de la tecnología

- Los alumnos utilizan la tecnología en muchas situaciones, como en su vida social, para el aprendizaje, y para escribir sus trabajos. Entonces, ¿por qué se les sigue exigiendo que respondan sus exámenes a mano?
- El IB utilizará la tecnología para ofrecer una evaluación de mayor calidad y más significativa. La tecnología no influirá en los enfoques de evaluación que se elijan.
- Es importante distinguir entre el efecto de la tecnología para respaldar a los examinadores expertos (corrección electrónica) y el uso de la tecnología para crear evaluaciones significativas para los alumnos (evaluaciones en pantalla —tanto para exámenes como para evaluaciones internas— y carpetas electrónicas).
- En pantalla no significa en línea: las evaluaciones del IB están diseñadas para que no sea necesario conectarse a Internet al realizarlas. Después de realizar la evaluación, sí será necesario utilizar Internet para enviar los archivos electrónicos al IB.
- Los alumnos deberán aprender en clase a utilizar la tecnología que se utilizará para realizar la evaluación. En general, será más fácil que se acostumbren a esta tecnología por medio de un método pedagógico que directamente en el momento de la evaluación.

## ¿Cómo es una evaluación en pantalla?

Una evaluación en pantalla consiste simplemente en realizar un examen en una computadora. La evaluación en pantalla no implica preguntas de opción múltiple o de arrastrar y soltar las respuestas (aunque puede incluir preguntas de este tipo si resulta adecuado). De hecho, se ha partido del estilo actual de los cuestionarios de examen, pero aplicado en una computadora. A partir de ahí, se empezarán a mejorar las evaluaciones; para ello, se incluirán preguntas interactivas que hagan que los exámenes resulten más auténticos o más accesibles para los alumnos. En el PAI ya se ha empezado a explorar el potencial que ofrecen las evaluaciones en pantalla, pero para el PD se ha decidido realizar una adopción más lenta. En el siguiente video se muestran ejemplos del tipo de preguntas que pueden ofrecer las evaluaciones en pantalla.

Evaluación electrónica del Programe de los Años Intermedios (PAI) del IB

En la práctica, la evaluación en pantalla supone que, en los exámenes, cada alumno deberá disponer de una computadora. Sin embargo, no será necesario que estén conectados a Internet, ya que los exámenes se habrán cargado en las computadoras previamente. Actualmente, en el PAI se utiliza el modelo siguiente: el software del examen "bloquea" la computadora mientras se ejecuta el examen y registra los momentos en que, por algún motivo, el software sufre alguna interrupción (el supervisor puede reiniciar la computadora en caso de interrupción justificada del examen).

La evaluación en pantalla ofrece ventajas significativas, ya que permite evaluar lo que realmente se quiere evaluar, y no simplemente lo que es posible con un examen en papel. Esto se describe más detalladamente en la sección sobre las ventajas para la validez, pero algunos ejemplos sencillos son poder ofrecer materiales en video, permitir que los alumnos muevan y roten diagramas, y posibilitar el uso de herramientas de procesamiento de textos para escribir ensayos o respuestas largas. En la sección "Equidad para todos: cómo satisfacer las necesidades de los alumnos" se explica de qué manera la evaluación en pantalla puede contribuir a la accesibilidad, ya que permite a los alumnos controlar los colores y el tamaño de fuente, y también utilizar herramientas tales como lectores de pantalla.

También hay que tener en cuenta que la evaluación en pantalla no será algo estático, sino que se irá desarrollando a medida que avance la tecnología. Actualmente, en este tipo de evaluaciones los alumnos



solamente utilizan 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.

#### La enseñanza mediante el uso de la tecnología

Actualmente, la tecnología está cambiando la manera en la que se enseña a los alumnos: ahora las aulas cuentan con pizarras digitales interactivas, y hay cursos en línea masivos y abiertos (MOOC, por sus siglas en inglés) que acercan la educación a un público cada vez más amplio. La mayoría de los alumnos no escribe sus trabajos a mano, sino utilizando una computadora.

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

Por otro lado, al introducir las evaluaciones en pantalla es importante asegurarse de que los alumnos estén familiarizados 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 de los alumnos se utilicen las mismas herramientas que en la evaluación.

### ¿Qué es la corrección electrónica?

Corrección electrónica es el término que se utiliza para designar el sistema de corrección en el que se usan computadoras para visualizar las respuestas de los alumnos y para registrar las puntuaciones. En general, los examinadores ya no reciben en papel ni los exámenes ni los trabajos de evaluación interna: ahora estos componentes se escanean para convertirlos a formato electrónico, de manera que los examinadores puedan acceder a ellos por Internet utilizando un software de corrección especializado.

Figura 3

Varios examinadores del IB utilizan el software de corrección electrónica para revisar trabajos de alumnos

durante una reunión de concesión de calificaciones.



Este cambio supone una serie de grandes ventajas que favorecen la velocidad y la calidad de la corrección:

- El IB siempre dispone del trabajo del alumno; si por algún motivo un segundo examinador tiene que revisarlo, el trabajo está disponible inmediatamente en vez de tener que enviarse por correo postal a otra parte del mundo.
- El trabajo del alumno puede anonimizarse para reducir la posibilidad de que haya sesgos en la corrección que realiza el examinador.
- Se 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 de los alumnos de un colegio pueden repartirse aleatoriamente entre todos los examinadores, en vez de que haya un único examinador 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 los alumnos entreguen su trabajo digitalmente. Actualmente, la mayoría de los exámenes se contestan a mano y posteriormente se escanean para convertirlos a formato digital.

## Riesgos de la evaluación en pantalla que se deben evitar

Desarrollar la evaluación en pantalla requiere ser consciente de una serie de desafíos que se deben superar para garantizar un sistema equitativo para todos los alumnos y los colegios. A continuación se presenta una lista de algunos de los riesgos clave que se gestionan. 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 electrónica.

#### Carga para los colegios

Hay que reconocer que, actualmente, el uso de la tecnología forma parte del mundo de los alumnos. Por ese motivo, resultaría anacrónico exigirles que escriban en papel. Sin embargo, también está claro que los colegios necesitan tiempo para poder cambiar sus procesos, y que las evaluaciones en pantalla requieren unas disposiciones distintas que los exámenes en papel. Por ello, las evaluaciones en pantalla del PD y del POP se irán introduciendo de manera equilibrada, con el fin de 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.

#### Riesgo de fallos en la tecnología

Cuando los alumnos realizan sus exámenes, resulta inadmisible que haya cualquier tipo de fallo tecnológico. Todos los alumnos deben disfrutar de una experiencia tranquila y sin interrupciones. Para lograr esto, el IB sigue dos principios. En primer lugar, la evaluación se debe realizar en una sola computadora que no necesite conexión a Internet. En segundo lugar, hay un proceso exhaustivo de puesta a prueba y cumplimiento de requisitos que los colegios tienen a su disposición el día antes del examen. La experiencia indica que la mayoría de los problemas que surgen se deben a la infraestructura o a determinados parámetros del colegio, y estos problemas pueden identificarse y resolverse con suficiente antelación a los exámenes.

#### Cuestiones de seguridad

Las evaluaciones en pantalla deben ser al menos igual de seguras que los exámenes en papel. Estas evaluaciones están diseñadas para que, mientras el alumno las está realizando, el resto de funcionalidades de la computadora queden bloqueadas. Además, las evaluaciones están cifradas y protegidas por 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.



#### El 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 tanto, el IB seguirá informando sobre cómo añade valor a la validez de sus evaluaciones, y el proceso de desarrollo de exámenes reflejará este valor añadido para evitar un uso indiscriminado de la tecnología.

#### Sesgos hacia grupos concretos de alumnos

Reconocemos que el uso de las computadoras para realizar las evaluaciones provocará que algunos alumnos tengan mejores resultados que otros. Sin embargo, no hay que olvidar que los exámenes en papel también se ven afectados por algunos sesgos: por ejemplo, los alumnos a los que les resulta difícil escribir con un bolígrafo durante mucho tiempo o que son lentos escribiendo a mano están en desventaja.

En todo caso, en las evaluaciones tanto en papel como en pantalla, el IB parte del principio de que no se intenta medir la capacidad del alumno de utilizar un bolígrafo ni una computadora. Por eso, las tareas de evaluación se diseñarán de tal modo que no favorezcan a aquellos alumnos que escriben rápido, ya sea a mano o en una computadora.

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, y este significado se protegerá mediante el criterio experto de los examinadores supervisores, que se basa en los datos acerca de los resultados. Por ejemplo, puede resultar más sencillo realizar una redacción bien estructurada si el alumno 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 que se sientan igual de cómodos trabajando en papel y en pantalla obtengan los mismos resultados en lo que resulta importante: reflejar su comprensión y su capacidad analítica en la asignatura.

La evaluación en pantalla permite comprobar nuevas características, así como características que el IB siempre ha valorado pero que anteriormente no se podían evaluar. Por ese motivo, es posible que haya que cambiar las expectativas de lo que se espera que los alumnos demuestren en las evaluaciones (véase la sección " Definición de los 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.

#### Obstáculos para los colegios que ofrecen programas del IB

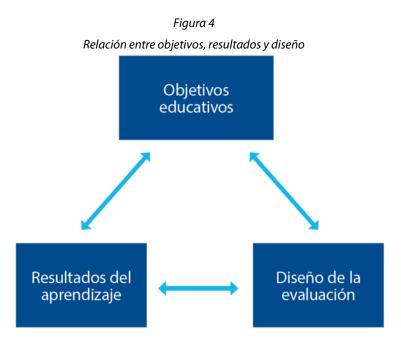
Volviendo al punto de la carga para los colegios, se trabajará con la comunidad del IB para establecer plazos que permitan que los colegios que lo deseen puedan aprovechar las oportunidades que ofrece la tecnología para realizar evaluaciones de mejor calidad. El IB considera que las posibilidades que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

## ¿Qué son los 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, y el más importante es que las evaluaciones deben servir para respaldar la educación, no para distorsionarla.

Las **prácticas de evaluación** son la manera en la que se aplican los 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.

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



## Terminología de la evaluación

- Tal y como sucede en la mayoría de las áreas, la evaluación dispone de su propia terminología y hay varios términos que resulta útil conocer al hablar sobre el tema.
- La introducción de la evaluación en pantalla provocará que haya que realizar algunos cambios terminológicos.

## Términos clave que es necesario comprender

Al igual que sucede en muchas esferas de la vida, incluida la enseñanza, la evaluación dispone de su propia terminología y vocabulario, lo cual puede dificultar entablar una discusión sobre la evaluación. Por eso, el IB está comprometido a utilizar un lenguaje sencillo y a reducir el uso de terminología específica.

A pesar de este esfuerzo de simplificación, es inevitable emplear determinados vocablos y conceptos específicos. A continuación se presentan algunos que resulta útil entender antes de leer este documento.

## Alumno matriculado y alumno

Un alumno matriculado es alguien que realiza una evaluación del IB. Un alumno es alguien que estudia los cursos del IB. Los alumnos pasan a ser alumnos matriculados cuando realizan la evaluación.

En general, esta diferencia terminológica solo resulta importante cuando se trabaja con datos en los que los números o los resultados puedan ser distintos entre los dos grupos.

#### Evaluación, examen, prueba y componente

La evaluación es cualquiera de las tareas que completa un alumno matriculado para demostrar sus capacidades. Un examen es una forma de evaluación en la que los alumnos responden, en condiciones estrictamente controladas, una serie de preguntas que formula el IB.

En este documento, la palabra prueba se utiliza genéricamente como sinónimo de examen, 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 se divide en partes que se realizan en momentos distintos. Estas partes se conocen como componentes de la evaluación, o simplemente componentes. Algunos ejemplos de componentes son los exámenes orales, la prueba 1 (de un examen), la prueba 2 (de un examen), o la evaluación interna.

## Puntuación y calificación

Las puntuaciones se otorgan para reflejar en qué medida el alumno ha respondido correctamente a una pregunta. La asignación de puntos varía en función de las preguntas y de los exámenes. En cambio, una calificación describe la calidad del desempeño del alumno, y debe significar lo mismo en todos los exámenes, cursos y asignaturas. Es muy importante distinguir estos dos conceptos.

Para obtener más información sobre este tema, consulte la sección "¿Cuál es la diferencia entre puntuaciones y calificaciones?".

### Pregunta y elemento

Se utiliza *pregunta* como término general para una tarea concreta; es posible que dicha tarea se estructure en distintas partes que guíen al alumno para tratar un problema.

Un *elemento* de la pregunta o subpregunta es cada respuesta individual que da un alumno. Por ejemplo, la pregunta 1, parte c (1c) o la pregunta 7, parte e, sección iii (7(e)(iii)) serían elementos.

### Autor del examen y examinador

El *autor del examen* es la persona que crea las preguntas y el esquema de calificación asociado que se utilizará para las evaluaciones. Casi siempre se trata de examinadores supervisores, si bien este trabajo requiere una serie de habilidades distintas y no está necesariamente relacionado con el papel del examinador.

El *examinador* es la persona que corrige el trabajo del alumno. En los equipos de examinadores existe una jerarquía, y los dos cargos más importantes son el de examinador principal (para cada componente) y el de examinador jefe (para una asignatura).

Los examinadores supervisores, bajo la dirección de sus examinadores principales y examinadores jefe, son los responsables de recomendar dónde se deben establecer los límites de calificación. Estos límites se deben establecer después de que se finalicen todas las correcciones.

Para obtener más información sobre las responsabilidades de los distintos cargos, consulte la sección "El proceso de evaluación: funciones y responsabilidades".



## Nuevos términos relacionados con la evaluación electrónica

El paso hacia la evaluación electrónica implica que el IB y la comunidad evaluativa, en su sentido amplio, deben empezar a utilizar un nuevo vocabulario para describir lo que hacen. A veces se recurre a añadir la palabra electrónico tras una palabra que ya se utilizaba, como es el caso de corrección electrónica, que es el término empleado para referirse a la versión electrónica de una corrección. Sin embargo, hay otros términos que difieren mucho de su equivalente para la versión en papel.

En la lista que se presenta a continuación, el único término que se utiliza ampliamente en este documento es evaluación en pantalla, aunque es posible encontrar los otros términos si lee otros materiales relacionados con la evaluación electrónica del IB.

#### Evaluación en pantalla

La evaluación en pantalla se refiere a cualquier evaluación que se realice utilizando una computadora. Esto engloba la idea de los exámenes que se contestan mediante una computadora, pero también puede incluir el trabajo de evaluación que se hace en clase si se realiza principalmente en un entorno informático.

La diferencia clave entre una evaluación en pantalla y la utilización de una computadora en clase es que, en la primera, se utiliza un software estrictamente controlado que se encarga de registrar las respuestas de los alumnos. Por ejemplo, es posible que se permita a los alumnos utilizar Internet para investigar sobre un tema, pero en la evaluación en pantalla es probable que el software les guíe en su búsqueda o registre los sitios que ha visitado el alumno para así poder evaluar la manera en que navega por Internet.

El término en pantalla se ha elegido cuidadosamente. La evaluación no se realiza en línea y, en concreto, los alumnos no deben estar conectados a Internet mientras realizan los exámenes en pantalla del IB.

### Archivo de respuestas y respuesta del alumno

En el mundo en papel, se habla de los exámenes escritos del alumno, pero al empezar a utilizar la amplia variedad de oportunidades que ofrece la evaluación electrónica, el trabajo también podría ser multimedia. El proceso de crear el trabajo forma parte de lo que se evalúa. En consecuencia, el producto de una evaluación electrónica se denomina respuesta del alumno.

El archivo de respuestas es simplemente el archivo informático que contiene las respuestas que ha proporcionado el alumno. Ese es el archivo que se debe enviar al IB.

#### Herramienta de familiarización

Resulta fundamental que los alumnos sepan utilizar las computadoras de manera eficaz para no encontrarse en desventaja a la hora de realizar sus exámenes. La herramienta de familiarización es un ejemplo en blanco de una evaluación en pantalla donde los alumnos pueden practicar el uso de las distintas funciones y características de la evaluación.



Figura 5 Captura de pantalla correspondiente a la herramienta de familiarización del PAI

Los alumnos también deberán realizar simulacros de examen utilizando la evaluación en pantalla, con límite de tiempo y esforzándose en sus respuestas a las preguntas. Por el contrario, para habituarse con el funcionamiento del *software*, los alumnos pueden disponer de acceso sin restricciones a la herramienta de familiarización.

### En caso de duda sobre algún término

Está claro que en las evaluaciones del IB se utilizan abundantes términos con los que tal vez los profesores y los alumnos no estén familiarizados. En todos los documentos del IB se procura mantener la precisión del contenido y evitar confusiones.

Como parte de este compromiso, el IB limitará el uso de abreviaturas y se cerciorará de que todas las siglas aparezcan desarrolladas en su término completo la primera vez que se utilicen en cada sección. Asimismo, los términos importantes se explicarán la primera vez que se utilicen, ya sea en el propio documento o en enlaces a las correspondientes explicaciones.

Si encuentra alguna palabra que no le resulte conocida, este documento ofrece un glosario con una lista completa de términos relacionados con la evaluación.



## ¿En qué consiste la evaluación?

- La evaluación puede significar muchas cosas distintas.
- 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".
- Es importante diseñar la evaluación cuidadosamente con el fin de alcanzar los propósitos para los que se utilizan sus resultados. Por ejemplo, una buena prueba formativa puede resultar muy deficiente para hacer una medición de carácter sumativo.
- La evaluación puede influir en las prácticas de enseñanza, y se debe diseñar para que el "efecto de repercusión" sea positivo.

### ¿Qué es la evaluación?

El término evaluación puede referirse a los distintos modos en que se pueden obtener y sopesar datos respecto al desempeño de un alumno. 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. En ocasiones el trabajo lo juzga el profesor del alumno y, en otros casos, lo evalúa un examinador externo.

Seguramente se habrá dado cuenta de que se han utilizado los términos juzgar y evaluar, en vez de corregir o calificar. Esto se debe a que hay una diferencia importante entre estos dos conceptos, que se explicará más adelante.



Figura 6 Ejemplos de distintos tipos de evaluación

Las tareas de evaluación que comprueban el nivel de competencia de un alumno pueden utilizar un modelo de **compensación** o un modelo de **dominio**. El modelo de compensación se utiliza en la mayoría de exámenes externos y permite demostrar un desempeño excelente en un área para mitigar un desempeño más bajo en otra. Por ejemplo, pensemos en un examen en el que hay 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 que obtuviese unas puntuaciones de 5, 5 y 5 (la mitad de la puntuación máxima para cada pregunta) o un alumno que obtuviese 10, 3 y 2 puntos respectivamente. En el segundo caso, una puntuación perfecta en la primera pregunta compensa las bajas puntuaciones logradas en las otras dos preguntas.

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. En el ejemplo anterior, si fuera necesario obtener 5 puntos de cada pregunta, el primer alumno aprobaría, pero el segundo no. Lo mismo ocurriría con un tercer alumno que obtuviera 10, 10 y 4 puntos: pese a alcanzar una puntuación total de 24, al no obtener los 5 puntos necesarios en la tercera pregunta, este alumno tampoco aprobaría.

Comparación entre el modelo de dominio y el modelo de compensación NO APROBADO **APROBADO** La media general supera No satisfactorio en la puntuación de aprobado el criterio 2 de 3 sobre 5 5 5 3 3 3 -3 3 3 3

Figura 7

Comparación entre el modelo de dominio y el modelo de compensación

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 un vestido, ser muy bueno en diseño no puede compensar el hecho de no saber coser.

Modelo de compensación

El IB utiliza varias herramientas de evaluación, entre ellas los cuestionarios de examen que se deben realizar al final del programa, y muchas otras tareas en distintas asignaturas que los alumnos completan 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.

# Evaluación formativa, sumativa y como aprendizaje: ¿por qué se evalúa?

Modelo de dominio

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 motivos para realizar las evaluaciones: uno es formativo y el otro es sumativo.

En el caso de la evaluación formativa, el objetivo es ofrecer comentarios detallados a los profesores y a los alumnos sobre los puntos fuertes y débiles de los alumnos, para así ayudarles a desarrollar sus capacidades.

En este sentido, los tipos de evaluación que implican una interacción directa entre el profesor y el alumno (como puede ser una discusión) resultan especialmente útiles.

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

Este concepto es similar a la noción de "andamiaje" de Wood et al. (1976), según la cual el profesor proporciona el andamio para construir el aprendizaje, pero solo el alumno puede llevar a cabo dicha construcción. El objetivo del profesor debe ser establecer evaluaciones formativas que planteen a los alumnos desafíos de un nivel adecuado, y continuar ajustando dicho nivel a medida que los alumnos progresan.

Por su parte, la evaluación sumativa se centra en medir qué sabe hacer el alumno, normalmente para demostrar que ha completado un programa de capacitación o que está preparado para pasar a la siguiente etapa de la educación que recibe. Es decir, la evaluación formativa está interesada en por qué el alumno hace algo, mientras que en la evaluación sumativa lo importante es saber si el alumno 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 particular de la evaluación sumativa es emitir un juicio sobre el alumno, y no fundamentar la futura enseñanza. Para aquellos que no estén convencidos de la necesidad de la evaluación sumativa, se recomienda leer la sección "La descripción del éxito: logros del alumno en la evaluación sumativa".

Dos posibles diferencias de cómo se utilizan las evaluaciones formativas y sumativas El alumno Enseñanza realiza la Enseñanza y aprendizaje evaluación aprendizaje • El alumno realiza la evaluación · El resultado se utiliza para la selección El profesor identifica los puntos fuertes y débiles Evaluación sumativa **Evaluación formativa** 

Figura 8

En la evaluación formativa es más importante identificar correctamente el conocimiento, las habilidades y la comprensión que los alumnos aún no han desarrollado, en vez de medir con precisión el nivel de logro de cada alumno. El equilibrio entre estos conceptos 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 y la calidad de los comentarios: los resultados de la evaluación se utilizan para tomar decisiones sobre el alumno, con frecuencia relativas a una selección competitiva para oportunidades laborales o educativas, pero también para apoyar la futura enseñanza.

Hay que tener en cuenta que 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.

## El "efecto de repercusión" y el aprendizaje

Nadie creció porque le midieran.

Meighan, 2004.

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. Surgenor (2010) resume claramente por qué esto puede conllevar implicaciones negativas y también positivas.

En su obra de 1971, Snyder propuso que los alumnos crearan su propia comprensión del currículo basándose en mensajes implícitos y explícitos sobre los elementos que cuentan en la evaluación, lo cual se describe como "currículo oculto". Esto puede ayudar a entender cómo aprobar una asignatura, pero no sirve para comprender la asignatura. La cita que aparece a continuación lo resume con mucha claridad:

Cuando volví a realizar el examen, me centré solamente en aprobarlo. Obtuve una puntuación del 96 %, y aquel tipo no entendía por qué 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.

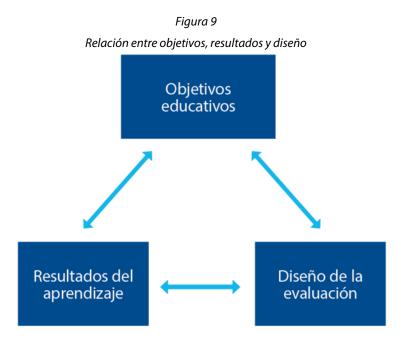
El concepto de *repercusión* también queda resumido en la frase tan oída de "Si no se va a evaluar, no se enseña", donde se destaca que la evaluación y la enseñanza no se pueden considerar como conceptos independientes entre sí.

La evaluación debería servir para aportar pruebas de que los alumnos han alcanzado los objetivos educativos, pero hay varias formas en que es posible que no lo logre.

Lo más probable es que los resultados del aprendizaje en los que se basa la evaluación no reflejen bien el propósito inicial de la educación. Un ejemplo en el terreno de la formación profesional sería un curso 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 cubre todos los resultados propuestos, la mayoría de las pruebas se centran en tareas que resulta fácil evaluar, pero que no tienen mucha importancia para los objetivos educativos generales.

Esta interdependencia de la evaluación y el propósito educativo se puede observar en el diagrama que aparece a continuación (del paradigma de Furst, 1958, 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.





Tal como se explica en el documento ¿Qué es la educación del IB?, el IB sigue un enfoque constructivista que le ofrece al alumno 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 una enseñanza y un aprendizaje eficaces, entonces se debe diseñar basándose en esta teoría constructivista del aprendizaje. Si desea seguir investigando sobre este concepto, puede consultar las obras de Black (1999), Shepard (1992), Wood (1998), y Lambert y Lines (2000).

La evaluación formativa es la que está más directamente relacionada con la manera en que los alumnos aprenden, 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 de la enseñanza y el aprendizaje.

### Puntuación y calificación son conceptos distintos

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

- Al corregir, se reconoce la labor que ha realizado un alumno y se le otorgan unas puntuaciones según un esquema de calificación o un marco similar. Esto indica en qué medida el alumno realizó correctamente la tarea de evaluación. La puntuación en sí misma no significa nada más.
- A la hora de decidir la calificación, el examinador juzga la calidad del trabajo del alumno 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.

Figura 10 La puntuación sería la calidad (el sabor) de una comida que se ha cocinado, mientras que la calificación refleja la complejidad del plato que se quería cocinar.



Es posible que un alumno reciba una calificación alta al responder bien solamente una parte pequeña de una pregunta muy difícil, pero que no pueda conseguir la misma calificación si solo responde correctamente muchas preguntas triviales.

Tal como se indica en secciones posteriores, no es necesario hacer referencia a los logros del alumno 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 en comparación con sus compañeros.

En sus evaluaciones, el IB normalmente utiliza la puntuación como indicativo del desempeño general (modelo de compensación), y posteriormente estudia el desempeño de los alumnos con esa puntuación para determinar una puntuación límite (límite de calificación) en la que se concede una calificación concreta a los alumnos que superan una puntuación determinada. Este proceso se explica con más detalle en la sección "Las prácticas de evaluación del IB".



## En qué consiste ser adecuado al propósito: el concepto de validez

- Que una evaluación sea válida significa que es 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

## Significado de validez

Todo el mundo es un genio. Sin embargo, si se juzga a un pez por su habilidad para trepar árboles, vivirá toda su vida pensando que es estúpido.

Anónimo, pero a menudo atribuida a Albert Einstein.

¿Cuál es el propósito de realizar exámenes? ¿Para qué sirven? Estas preguntas aparentemente sencillas 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:

- Determinar qué comprende el alumno una vez concluido el programa de estudios
- Ser un medio de selección para estudios posteriores o para el mundo laboral
- Ser un indicador para saber si el alumno tendrá éxito en el futuro
- 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, ¿deberíamos centrarnos en el análisis? ¿Qué relación guarda esto con el primer objetivo de determinar qué sabe el alumno (en geometría)?

Estas cuatro posibilidades no son definitivas. Para Newton (2007), los resultados de la evaluación pueden tener diversos usos, y enumera una serie de ejemplos:

- Evaluación social
- Formativos
- Supervisión de los alumnos
- Transferencia
- Determinación de a qué curso asignar a un alumno
- Diagnóstico
- Orientación
- Medición de aptitudes

- Obtención o renovación de licencias
- Elección de colegio
- Supervisión de la institución
- Asignación de recursos
- Intervención organizativa
- Evaluación del programa
- Supervisión del sistema
- Equiparabilidad

#### • Selección

Dada esta variedad abrumadora de usos que entran en conflicto entre sí, resultaría tentador sugerir que la evaluación y las calificaciones de los exámenes no pueden ser adecuadas al propósito y, por lo tanto, hay que evitarlas. El problema de este enfoque es que sigue siendo necesario que las personas tomen decisiones y, si no tienen disponibles los resultados de la evaluación, se corre el riesgo de que utilicen otras maneras peor diseñadas y peor comprendidas para tomar dichas decisiones. A esto es a lo que se refiere Mike Cresswell en la siguiente cita:

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.

La noción de validez engloba el concepto de que el título académico sea adecuado al propósito.

El término validez se usa a menudo, pero hace mucho tiempo que los especialistas académicos 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 fiabilidad, para referirse a la capacidad de una evaluación para medir realmente lo que debe medir. Un ejemplo de esto sería no utilizar un examen escrito para demostrar que el alumno sabe hacer un pastel. Para evitar confusiones, en este documento nos referiremos a este segundo concepto como pertinencia del constructo. Para obtener más información, consulte la sección "Pertinencia del constructo y autenticidad".

En última instancia, la validez es una cuestión de juicio. Las demandas opuestas de una evaluación a las que nos hemos referido anteriormente se explicarán con más detalle más adelante, pero siempre habrá que llegar a un punto de equilibrio entre dichas demandas. Para resolver la cuestión, hay que plantearse si las pruebas disponibles indican que la evaluación es lo suficientemente "adecuada al propósito" como para resultar útil. Por tanto, al hablar de validez es recomendable reflexionar sobre la solidez de las pruebas aportadas, es decir, del "argumento de validez" que se plantea, en lugar de buscar un sí o un no como respuesta inmediata. En la siguiente cita se cubre esta idea con más detalle:

Si la declaración de validez actúa como promesa o garantía que da a los usuarios luz verde para interpretar y utilizar los resultados de la evaluación según lo especificado, esto plantea un punto fundamental final que no siempre recibe la atención que merece: sin una formulación explícita de la afirmación de validez, los usuarios no podrán interpretar y utilizar los resultados de la evaluación de manera adecuada.

Newton, 2007

Por último, la validez no es algo que se logre o no 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, sino que se amplía continuamente y se va refinando durante el proceso.

### ¿Qué se considera válido: la evaluación o su uso?

Veamos la siguiente situación:

Un alumno lleva dos años estudiando la asignatura de Física y ha logrado una calificación de 6. El examen se realizó en español.

¿Esa calificación significa que el alumno domina el español? Si llevamos este ejemplo al límite del absurdo, ¿significa que el alumno cocina bien?

En este ejemplo, el examen que realizó el alumno puede haber sido una buena prueba de su comprensión de la física, pero se está utilizando como indicador de su comprensión del español. Por tanto, ¿resulta válida la evaluación?

Otro ejemplo más sutil podría ser plantearse si una calificación concreta en Matemáticas indica que el alumno es competente en aritmética (que es solamente un aspecto del currículo). Esto introduce la idea de que lo válido no es la evaluación (o el resultado), sino el propósito al que se somete. Newton (2012) lo expresa de la siguiente manera: "el uso de un procedimiento de evaluación concreto para un propósito específico (p. ej., tomar una decisión específica) resulta válido si su argumento interpretativo es suficientemente sólido".



Si lo válido es el propósito para el que se utiliza la evaluación, y no la evaluación en sí misma, ¿qué responsabilidad tiene el IB de demostrar que sus exámenes son válidos? ¿Es necesario mostrar la validez de cada uso posible? ¿Resulta adecuado definir una cantidad limitada de usos que el IB considera válidos? ¿Se deberían indicar explícitamente los usos que el IB no considere válidos?

Se necesita una solución práctica a este problema. Las evaluaciones se diseñan tomando en consideración un número específico de propósitos para los que el IB realizará un argumento de validez. También se tiene en cuenta el grado de conformidad con otros usos comunes que se encuentran en la filosofía del IB. Si el análisis de estas dos categorías de utilización muestra que los argumentos son suficientemente sólidos, se considerará de manera informal que la evaluación resulta válida.

### 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. Esto significa que no se puede "demostrar" la validez, sino que se debe preparar un argumento convincente sobre por qué las decisiones tomadas conducen a un curso y una evaluación significativos.

Por razones similares, las pruebas de los argumentos de validez deberían surgir de manera natural del proceso de desarrollar, supervisar e impartir un programa. Además, deberían reflejar las discusiones y la decisión alcanzadas. 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. Dichas preguntas se pueden consultar en el "Anexo 2: Hoja de ruta para crear un argumento de validez".

#### 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 también posteriormente, siempre y cuando 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 la evaluación es válida (es decir, el argumento 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, especialmente 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 equidad, la equiparabilidad, y la pertinencia del constructo. Los comentarios de los profesores y los resultados de los alumnos 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 posteriormente 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. En el siguiente diagrama se resume este proceso.

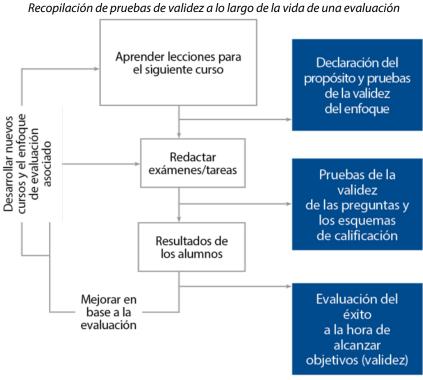


Figura 11
Recopilación de pruebas de validez a lo largo de la vida de una evaluación

Aunque el conjunto completo de pruebas para el argumento de validez solamente puede existir después de que los alumnos hayan realizado la evaluación, cada etapa ofrece una base para el siguiente paso del proceso.

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

Sireci, 2007.

# Para la validez, ¿cuáles son las ventajas de trabajar en pantalla? ¿Y las ventajas de la evaluación electrónica?

- En el mundo actual, el papel ya no es el soporte preponderante: las computadoras 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 los alumnos.
- La versatilidad de las computadoras permite a los alumnos hacer las modificaciones visuales y sonoras que necesiten para acceder a las preguntas, en vez de tener que solicitar exámenes especiales con varios meses de antelación.
- A pesar de las preocupaciones que puedan surgir, la evaluación electrónica permite protegerse mejor contra las conductas improcedentes y la mala administración.

Casi con toda probabilidad, estará leyendo este documento en una pantalla, ya que se trata de un recurso que solo está disponible electrónicamente. Los correos electrónicos, los mensajes de texto y los medios sociales son actualmente mucho más frecuentes que las cartas como medio de comunicación. Asimismo,

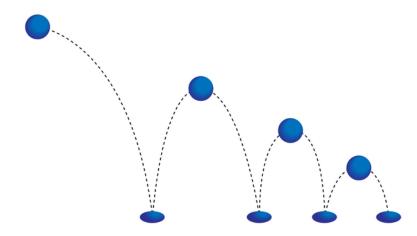


hoy en día muchos empleos requieren utilizar computadoras, por lo que ya no se trata de algo excepcional. Por lógica, ofrecer evaluaciones auténticas pasa por incluir en ellas el uso de las computadoras.

Tradicionalmente, las evaluaciones han resultado muy limitadas en lo referido a aquello que pueden cubrir. En general, en los exámenes escritos, los alumnos se han visto limitados 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 en pantalla ofrecen la posibilidad de incluir como estímulo elementos de video y audio que el propio alumno puede reproducir y pausar, en vez de tener que depender de un supervisor de examen que se encargue de reproducir el material para toda la clase. Además, este sistema permite utilizar diagramas animados donde resulte adecuado, en vez de incluir una imagen estática y una descripción larga sobre el movimiento de un objeto.

Figura 12

La claridad de una pregunta de la evaluación electrónica sobre la trayectoria de una pelota que está rebotando podría mejorarse incluyendo una imagen en movimiento en lugar de una imagen estática, tal como se muestra aquí.



A medida que las evaluaciones electrónicas evolucionan, se hace posible evaluar de qué manera el alumno interactúa con un problema, responde ante información nueva o desarrolla una simulación. En última instancia, la evaluación en pantalla debería permitir responder a las reacciones del alumno. Así se recrearían las oportunidades que ofrecen las evaluaciones que realizan los profesores, como los exámenes orales, pero sin los problemas que provoca el hecho de que distintos alumnos tengan profesores diferentes.

El beneficio clave de la evaluación electrónica es que permite evaluar lo que realmente se quiere medir, sin limitarse a lo que se puede juzgar mediante un examen en papel.

Otro aspecto de la validez es reducir los sesgos, sobre todo los relacionados con los alumnos con necesidades específicas de acceso a la evaluación. Periódicamente, el IB procesa solicitudes para realizar cuestionarios de examen con distintos tamaños de letra o en papel de otro color. Asimismo, también recibe solicitudes para preparar cuestionarios de examen en formato braille.

Si bien los exámenes en pantalla no pueden eliminar estas barreras, sí que permiten a los alumnos seleccionar el tamaño de letra y el color que mejor se adapten a sus necesidades. El uso de los lectores de pantalla también ofrece otros beneficios destacables en lo referido a la accesibilidad de los alumnos. Así, aunque es posible que algunos alumnos no puedan utilizar una computadora para cubrir sus necesidades específicas de acceso a la evaluación, la evaluación en pantalla permite que una cantidad mayor de alumnos pueda acceder a las pruebas sin necesitar apoyo adicional, si bien este apoyo se podrá seguir utilizando cuando resulte adecuado.



Figura 13 ¿Cuáles pueden ser las vulnerabilidades de las evaluaciones electrónicas?

Uno de los temas que más preocupación causa en cuanto a la evaluación electrónica es que resulte menos segura que los exámenes en papel. En realidad, las evaluaciones en pantalla están protegidas frente a algunos riesgos de los exámenes en papel, si bien se exponen a otros.

Una de las mayores ventajas de la evaluación electrónica es que se puede enviar electrónicamente a prácticamente cualquier lugar del mundo y permanecer bloqueada hasta que se proporcione la contraseña a los alumnos el día del examen. 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 electrónica resulta más ventajosa. Aunque se violase la seguridad del colegio, los ladrones aún tendrían que lograr acceder a las evaluaciones electrónicas, que se han diseñado para evitar que esto suceda. En todo caso, si finalmente los ladrones consiguieran acceder a las evaluaciones electrónicas, sería como si hubieran robado los exámenes en papel, pero dispondrían de menos tiempo para lograrlo, ya que las evaluaciones electrónicas se pueden enviar a los colegios cuando la fecha del examen esté más cercana.

El IB es consciente de que existe un riesgo durante el examen en pantalla, ya que los alumnos tienen acceso a una computadora en la que, en teoría, podrían conectarse para consultar apuntes no autorizados y otros recursos más fácilmente que con los métodos tradicionales de introducir a escondidas apuntes en la sala de examen. Sin embargo, las evaluaciones electrónicas del IB están diseñadas de manera que la computadora bloquea el acceso a cualquier otro programa mientras se está haciendo el examen. Además, la capacidad de registrar cómo responden a las preguntas los alumnos, así como sus respuestas finales, ofrece oportunidades para identificar comportamientos relacionados con conductas ilícitas.

En resumen, la evaluación electrónica supone nuevos retos en lo referido a la seguridad de los exámenes, pero al mismo tiempo elimina algunas de las oportunidades de cometer conductas ilícitas que existían en el pasado.



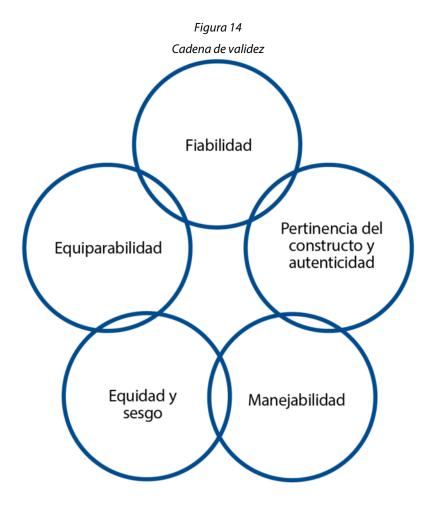
### Elementos de la cadena de validez

- La validez implica aspectos 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: la fiabilidad, la pertinencia del constructo, la manejabilidad, la equidad y la 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 aspecto.
- El IB pone mucho esmero en crear programas, cursos y evaluaciones cuyo constructo sea pertinente.
- En el ámbito de la evaluación, esto significa que se 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 (igual que la fiabilidad) se considera una característica esencial de cualquier sistema de evaluación, especialmente si es de alto impacto, en el que el resultado es muy importante para el alumno o el profesor. 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, Kane y Cohen, 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 alumno y el ingenio del colegio.

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



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 los alumnos. Otro ejemplo sería que se pretendiese contextualizar un examen para cada cultura o país específico (equidad), lo cual podría hacer surgir preguntas sobre si realmente se proporciona el mismo examen (equiparabilidad).

## El equilibrio entre los distintos aspectos de la validez

Al diseñar una evaluación, es importante hallar un equilibrio entre las distintas prioridades contrapuestas, 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 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 15
Hallar un equilibrio entre prioridades contrapuestas de varios aspectos de la validez: equidad y manejabilidad

Falta de equilibrio Carga poco razonable sobre los alumnos (manejabilidad)	Equilibrio correcto Se evalúa a los alumnos 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 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 del alumno se basa íntegramente en una tarea.

Las siguientes secciones mostrarán que, a pesar de que tienen definiciones diferentes, hay considerables coincidencias en la manera en que estos cinco elementos se dan y se gestionan. Estos elementos también son aspectos del concepto amplio de validez.

Por último, el IB da la mayor prioridad a la validez del 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 se define como "la medida en que el alumno obtendría el mismo resultado en una prueba si se repitiera el procedimiento empleado".

Como se explica a continuación, esto no es necesariamente lo mismo que decir que el alumno obtendría el "resultado correcto".

En su introducción al concepto de fiabilidad (Ofqual's Reliability Compendium, 2011), Winkley y Cresswell (2011) presentan una lista (no exhaustiva) de causas de una posible falta de fiabilidad:

- 1. Fiabilidad entre los distintos correctores. Un examinador puede ser más o menos tolerante que otro en determinadas preguntas (y puede incluso que el mismo corrector se muestre más o menos tolerante de un día para otro).
- 2. Variabilidad en el desempeño del alumno. Es posible que el desempeño de un alumno en un examen varíe un poco de un día a otro, sobre todo si cambian las condiciones en las que se realiza el examen (si es por la mañana o por la tarde, según quién sea la persona que administra el examen, si el alumno ha dormido bien la noche anterior, si desde el aula se oye a alguien cortando el césped, si el alumno tiene dolor de cabeza o no, etc.).

- 3. Distintos cuestionarios de examen. Las preguntas serán diferentes entre un cuestionario de examen y otro, y puede que evalúen diferentes facetas de la comprensión del alumno (normalmente los exámenes tratan sobre algunos temas del currículo, ya que no hay suficiente tiempo para evaluarlos todos, y es posible que los alumnos decidan estudiar algunos temas concretos y no otros).
- 4. Equiparabilidad de los resultados de un año a otro.
- 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.
- Distintos tipos de actividades de evaluación. Muchas evaluaciones están compuestas por diversos tipos de actividades, y esos métodos variados de evaluación pueden implicar diversos problemas de fiabilidad.
- 7. Distintos tipos de preguntas. Es posible que los alumnos obtengan diferentes resultados en función del tipo de preguntas que se les presenten.

En la práctica, el IB normalmente se plantea que el alumno solo realiza la evaluación una vez y busca la coherencia en el resultado del proceso, por lo que 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 la evaluación. 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 los examinadores realicen la corrección de manera coherente.

### Fiabilidad de la corrección

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

A continuación se incluyen cinco fragmentos de respuestas reales de una evaluación de Adquisición de Lenguas del IB. Al hacer clic en cada una de ellas, aparecerán los exámenes manuscritos de los alumnos.

Use el esquema de calificación que se muestra a continuación para evaluar sobre 10 la calidad de la lengua y registrar las puntuaciones que otorgaría a cada trabajo.

Una vez que haya "corregido" cada trabajo, vea los comentarios sobre la calidad de sus puntuaciones. Los comentarios aparecen después de los exámenes de los alumnos.

#### Esquema de calificación

¿En qué medida utiliza el alumno la lengua de forma correcta y eficaz?

Puntu ación	Descriptor de nivel
1–2	El manejo de la lengua es, en general, inadecuado.
	Se utiliza una variedad de vocabulario muy limitada, con muchos errores básicos.
	Las estructuras de las oraciones sencillas son muy pocas veces claras.
3-4	El manejo de la lengua es limitado y, en general, ineficaz.
	Se utiliza una variedad de vocabulario limitada, con muchos errores básicos.
	Las estructuras de las oraciones sencillas son a veces claras.
5-6	El manejo de la lengua es, en general, adecuado, a pesar de observarse muchas incorrecciones.
	Se utiliza una variedad de vocabulario un tanto limitada, con muchos errores.
	Las estructuras de las oraciones sencillas son en general claras.
7-8	El manejo de la lengua es eficaz, a pesar de observarse algunas incorrecciones.
	Se utiliza una variedad de vocabulario de forma correcta, con algunos errores.

Puntu ación	Descriptor de nivel
	Las estructuras de las oraciones sencillas son claras.
9-10	El manejo de la lengua es bueno y eficaz.
	Se utiliza una amplia variedad de vocabulario de forma correcta, con pocos errores de importancia.
	Las estructuras de algunas oraciones complejas son claras y eficaces.

Vea cada uno de los exámenes y asígneles una puntuación usando el esquema de calificación como guía.

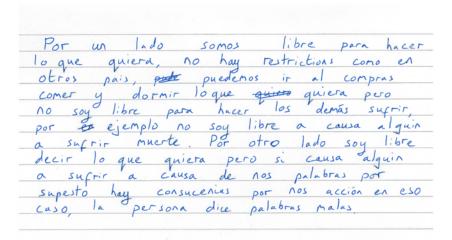
Figura 16

Ejercicio sobre la fiabilidad de la corrección

#### Alumno A

Como seres humanos, estamos libres de hacer lo que sea; cualquier cosa.
Se entiende que esto puede ser wisto 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 la que yo elijo haver podria tener consecuencias. Entonces
creo que todos ustedes saben la importancia de tener en
cuenta los sentimientos de tus compañeros antes de decid
algo, ¿verdad?

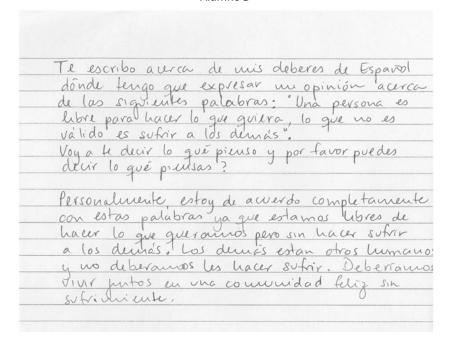
#### Alumno B



#### Alumno C

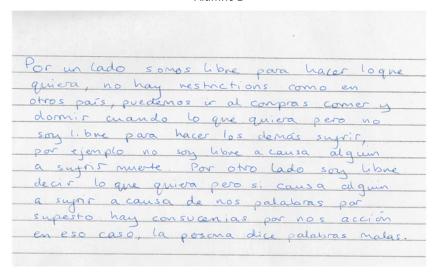
En prin	er lugar estoy o	le accuerdo o	on el facto qu	elas
	y las decisiones			
ser vali	tas si hace sufr	ir a los dema	S. Pero infeliz	mente
vivimos	a una sociedad	donde palabo	as puedu off	endir,
	hace sufrir y la			
son bas	tante sensitivas. t	tay veres done	se loque elm	udo

#### Alumno D





#### Alumno F



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

#### **Comentarios**

¿Ha observado que los trabajos de los alumnos B y E son idénticos, pero están escritos con una caligrafía diferente?

¿Ha otorgado la misma puntuación al alumno B que al alumno 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 de los alumnos, sino en poner de manifiesto que un mismo corrector 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).

#### Resultados coherentes no es lo mismo que resultados adecuados

Es importante entender que el objetivo de lograr un nivel alto de fiabilidad es que los alumnos obtengan la misma puntuación (justa) sea quien sea el examinador que corrija sus trabajos, y no una puntuación "adecuada". Para determinar la calidad de un trabajo, los docentes emplean su juicio profesional, pero con frecuencia cuesta que dos profesores se pongan de acuerdo sobre qué puntuación conceder. Se considera que hay fiabilidad cuando ambos se rigen por el mismo juicio o criterio (el del examinador supervisor).

Esto supone un desafío particular al abordar las consultas sobre los resultados. En este caso, el examinador 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.

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.

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.

Aproximadamente el 63 % de los profesores y los alumnos 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 correctores", 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, Opposs y Boyle, 2010.

## Pertinencia del constructo y autenticidad

¿Con qué precisión se mide aquello que se intenta medir? Las expresiones pertinencia del constructo y validez del constructo a veces utilizan indistintamente para referirse al mismo concepto. Sin embargo, para evitar que haya confusiones con el término general de *validez*, en este documento utilizaremos la expresión pertinencia del constructo.

La idea de una evaluación auténtica está estrechamente relacionada con la pertinencia del constructo. Esto significa que la prueba se realiza de manera que se ajuste a las situaciones en las que el alumno 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—, pero a menudo el enfoque 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. A los alumnos se les han enseñado los recursos literarios que se utilizan en un texto concreto, y después la evaluación les pide que redacten un ensayo relacionado con uno de esos recursos. Si los alumnos se limitan a recordar y reproducir todo aquello que les explicó el profesor, ¿realmente están demostrando su comprensión de la literatura?

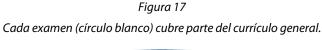
Comprender claramente qué se quiere que una tarea de evaluación permita demostrar al alumno, y analizar meticulosamente en qué medida le ha permitido demostrarlo, representan las mejores prácticas a fin de garantizar unos niveles adecuados de pertinencia del constructo. En particular, hay que plantearse qué otras habilidades necesitará el alumno para poder llevar a cabo la tarea. Las tareas abiertas, como los ensayos o proyectos, resultan especialmente proclives a exigir que el alumno demuestre una capacidad muy buena de redacción para poder demostrar las habilidades de investigación o de análisis que la tarea pretendía abordar. Esto queda englobado en el concepto de diseño universal de evaluación, que pretende reducir las barreras a los alumnos en los casos en que las dificultades no estén relacionadas con lo que la evaluación debe evaluar ( Dolan, Burling, Harms, Strain-Seymour, Way y Rose, 2013).

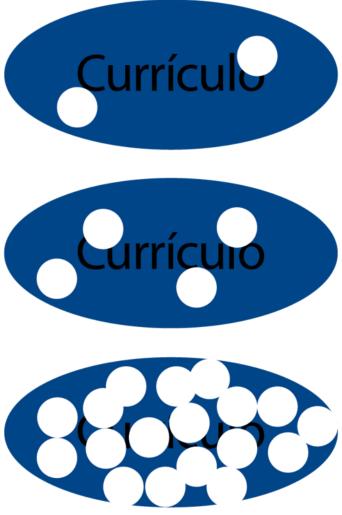
Históricamente, la forma de medir la pertinencia del constructo (o la validez del constructo) se basaba en gran parte en la psicometría. Esto llevó a un círculo vicioso en el que aquello que se pretendía evaluar con la prueba (es decir, el constructo) venía definido por la propia prueba. Si todas las preguntas de la prueba ofrecían un resultado coherente, se consideraba demostrado que la prueba era pertinente al constructo. Sin embargo, si una pregunta no reflejaba el resultado de las demás preguntas, se consideraba que no era pertinente al constructo y, por lo tanto, se descartaba. Los enfoques modernos tienden a evitar esta definición autorreferencial y puramente estadística de pertinencia del constructo.

La pertinencia del constructo está estrechamente vinculada con el currículo, tal como se indica en la figura 2 llamada "Relación entre objetivos, resultados y diseño". Muchas veces, una evaluación solamente planteará preguntas basadas en una selección reducida del material, y la pregunta sobre la pertinencia del constructo será considerar si se ha cubierto el currículo de manera suficiente. Del mismo modo, elegir los



tipos de preguntas que se van a plantear es una parte importante de este diseño, ya que los distintos tipos de preguntas o tareas tienen distintos puntos fuertes y débiles en lo que se refiere al tipo de constructo que pueden evaluar.





## Manejabilidad

A diferencia de los dos puntos anteriores, el aspecto de la manejabilidad no se ha investigado a fondo en la bibliografía especializada, 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. Por ello, resulta razonable abordar la manejabilidad de una evaluación desde el punto de vista del alumno, del colegio e incluso del IB.

Para los alumnos, la manejabilidad suele estar relacionada con el tiempo que dure la evaluación. Por ejemplo, se considera que un examen de 8 horas de duración no resulta razonable para una persona de 18 años. También se debe tener en cuenta en qué momento se realiza la evaluación. Por ejemplo, tal vez sea razonable hacer una serie de cuatro exámenes de una hora de duración cada uno, pero realizarlos el mismo día con tan solo una breve pausa entre ellos puede ser una carga muy pesada. El contexto del IB implica tener en cuenta la carga que suponen otras evaluaciones del IB que es posible que los alumnos realicen

como parte de su educación. Esto está relacionado con el principio de que se debe considerar la validez de los programas, y no solo de las asignaturas individuales.

Desde la perspectiva de un colegio, la manejabilidad también puede incluir los requisitos de proporcionar materiales para la evaluación. Así, por ejemplo, en un curso de formación profesional de ingeniería puede que sea necesario que cada alumno disponga de un motor que pueda montar. En el contexto de las titulaciones del IB, el requisito de poder realizar la evaluación en una computadora o un dispositivo similar conlleva plantearse si la evaluación sería manejable.

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

Por último, la manejabilidad también afecta al IB en lo que se refiere a la cantidad de trabajos de alumnos 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, esta no es una opción práctica si la tarea se tiene que revisar externamente. Por dar otro ejemplo, no sería razonable ofrecer una amplia variedad de preguntas opcionales si posteriormente 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. Si se aumenta la cantidad (duración) de la evaluación, se podrá demostrar mejor la comprensión del currículo completo por parte del alumno, y también aumentará la probabilidad de que las decisiones demasiado severas o demasiado generosas se acaben anulando las unas a las otras y queden compensadas. Sin embargo, dicho aumento también provocará que la evaluación mida la capacidad del alumno para mantener su desempeño en un examen largo, en vez de los objetivos de la evaluación.

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

## Equidad y sesgo

Se considera que una prueba está sesgada cuando confiere una ventaja a una persona sin que ese sea el objetivo de la evaluación. Para poner esto en perspectiva, podemos plantear los siguientes ejemplos:

- 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 (de pintura) en el que el caballete se coloca a dos metros del suelo.

En todos estos casos, las tareas (preguntas) que se plantean pueden representar una prueba razonable, pero algunos alumnos (los más bajos en el último ejemplo) se verán en clara desventaja por la situación. En la práctica, la mayoría de los 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 de los alumnos. Contextualizar las preguntas es especialmente complicado, ya que algunas situaciones pueden resultarles familiares a algunos alumnos pero no a otros, 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 en la aptitud o 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 preguntas o 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 algunos alumnos. 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, como pueden ser los períodos de temperaturas o de contaminación extremas en algunas partes del mundo, o durante períodos en los que los alumnos no pueden concentrarse o prepararse de manera adecuada. Esto supone un desafío especial para el IB, ya que cada región del mundo en la que el IB está presente tiene sus peculiaridades y requisitos, a menudo en conflicto con los de otros países, por lo que cualquier fecha que se elija supone problemas para unos colegios u otros, en función de donde se encuentren.

La distribución de la sala donde se realiza el examen puede constituir también un sesgo. Pongamos el ejemplo de dos alumnos 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 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 surgidos de la corrección

Para empezar esta sección, es importante destacar que la mayoría de los 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. Esto no significa que el IB no tenga el deber de mitigar estos sesgos, sino que 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 alumno (por ejemplo, Hughes et al., 1983), el trato preferente en función del género del alumno (en caso de que pueda saberlo o sospecharlo el evaluador), 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 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 de los alumnos 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 desarrolla una opinión positiva sobre un alumno 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.

#### Sesgos relacionados con las preguntas de evaluación

No hay sesgo cuando la diferencia tiene que ver con lo que realmente se está intentando medir. Por ejemplo, la diferencia de altura entre hombres y mujeres no se debe a que haya sesgos en el método que se utiliza para medir la altura. Aunque la evaluación no esté sesgada, es posible que sea discriminatoria.

El sesgo que surge de las propias tareas de evaluación es el problema más importante. En la construcción 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 (funcionamiento diferencial del elemento), entonces se considera sesgado y se elimina de la prueba. Los subgrupos de alumnos pueden definirse por género, etnia, clase social o competencia lingüística; de hecho, por cualquier característica definitoria que pueda considerarse no pertinente con respecto al constructo que es objeto de la prueba.

Las afirmaciones de la existencia de sesgos a favor o en contra de determinados subgrupos de alumnos no siempre son fáciles de justificar. En los años iniciales del desarrollo de las pruebas de inteligencia, se excluyeron aquellos elementos que daban lugar a una diferencia significativa de respuesta entre ambos géneros. Esto se basaba en el principio de que no debía existir ninguna diferencia de inteligencia entre varones y mujeres; por tanto, los elementos que revelasen una diferencia de ese tipo debían estar midiendo algo sin pertinencia para el caso. Ese punto de vista es, cuando menos, debatible, y varios autores han explicado que las diferencias de inteligencia entre diferentes grupos de personas están relacionadas con factores biológicos, del entorno o socioeconómicos, así como con la naturaleza de las propias pruebas.

En el desarrollo de las pruebas de inteligencia se ha puesto de manifiesto mayor preocupación por la fiabilidad de la medición que por la naturaleza de lo que de hecho se está midiendo, lo cual se ha moldeado para adaptarse a las exigencias de un alto grado de fiabilidad. Pueden existir aspectos importantes de un constructo que estén legítimamente ligados a ciertas características de grupos dentro del conjunto del alumnado. Las implicaciones de estas diferencias para la enseñanza y el aprendizaje son significativas. Un enfoque más adecuado es incluir en las pruebas una variedad equilibrada de elementos que den lugar a un rendimiento diferencial por parte de los diferentes subgrupos del alumnado, de modo que ningún subgrupo se encuentre en desventaja. Cualquiera que sea el enfoque que se adopte, puede establecer limitaciones considerables a la hora de diseñar una prueba.

Para determinar si existe sesgo en una pregunta o un elemento de una prueba, se debe plantear con cuidado cómo se puede vincular explícitamente la tarea con el constructo subyacente y cuáles podrían ser los posibles factores que introducirían 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 distintos.

- El contexto en el que se plantea la pregunta (por ejemplo, las referencias culturales a Estados Unidos benefician a los alumnos que se encuentran en ese país, y pueden desfavorecer a los alumnos que viven en otros sitios).
- Las preguntas de desarrollo sobre las relaciones humanas favorecen a los grupos en cuyas sociedades se fomenta la expresión de las emociones.
- Las preguntas de opción múltiple pueden favorecer a los chicos.
- Los trabajos de clase o los trabajos de proyecto pueden favorecer a las chicas.
- Una pregunta que utilice lenguaje o convenciones de una clase social favorecerá a los alumnos de dicha clase.
- Algunas preguntas pueden ser inteligibles solamente en el seno de ciertas culturas. Por ejemplo, una
  pregunta sobre ancianos que vivan solos puede resultar bastante ajena a ciertas culturas, o una
  pregunta que trate de un papel típicamente masculino o femenino en una cultura puede parecer muy
  fuera de lugar en otra.

En el caso del sesgo de género, existen pruebas de su existencia, sobre todo en los Estados Unidos, pero no queda claro qué aspectos del formato contribuyen a estos resultados.

Al diseñar las evaluaciones, no siempre está claro cómo se debería responder a estas diferencias. 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 los diferentes subgrupos 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 construir preguntas, y 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 causa preocupación saber cuántos subgrupos diferenciados del alumnado deben tenerse en cuenta. ¿Se debería tener en cuenta a los alumnos que tienen distintos estilos de aprendizaje, o a aquellos cuyo temperamento les impide realizar pruebas o exámenes formales? 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 que para otros. Al final, la preocupación sobre la posibilidad de sesgo en diferentes tipos de



preguntas y contextos a menudo se reduce a una cuestión sociopolítica. La equidad en la evaluación, que incluye evitar el sesgo, es un tema de suma importancia, especialmente en ciertos países donde cualquier sesgo demostrable en un instrumento de evaluación puede incluso 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) concluyeron 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 en su mano 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 pruebas de opción múltiple en los exámenes de inglés para mejorar el desempeño de los varones, ya que esto distorsionaría la validez de la evaluación.

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. Que diferentes subgrupos tengan un desempeño distinto en una prueba puede ser el resultado de factores que tengan bastante poco que ver con la propia prueba. Hay muchos otros factores causantes de falta de equidad en la educación que tienen un impacto decisivo en los logros del alumno, 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, y las diferencias en las circunstancias sociales y en el nivel de apoyo familiar que recibe cada alumno. Todos estos factores pueden influir de forma significativa en las perspectivas individuales de éxito educativo 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 la etnia de los alumnos, y señalaron que los intentos de ajustar los instrumentos de evaluación para remediar diferencias en el rendimiento por parte de diferentes grupos étnicos 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 de manera objetiva independientemente del contexto social y la cultura. El concepto de éxito en la educación se define y se mide en cualquier sociedad según los estándares de un pequeño sector de dicha sociedad.

#### Eliminación de sesgos para los alumnos con necesidades específicas de acceso a la evaluación

Otro aspecto del sesgo que debe combatirse es la posibilidad de que una tarea de evaluación discrimine negativa e injustamente a los alumnos con necesidades de apoyo para el aprendizaje debidas a motivos tales como dislexia, síndrome de déficit de atención, o problemas de visión. Para ello, hay que asegurarse de que las condiciones en las que se realizan las tareas de evaluación hagan las concesiones apropiadas para estos alumnos, de modo que puedan demostrar su nivel de logro educativo en condiciones de igualdad con respecto a los demás alumnos.

Este tema se trata con más detalle en la sección "Equidad para todos: cómo satisfacer las necesidades de los alumnos".

#### Cómo reconocer el sesgo

Es importante tener en cuenta que los sesgos pueden ser positivos y también negativos. Si un grupo de alumnos 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 todos los alumnos.

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. Por tanto, es necesario buscar pruebas de dichos sesgos en los resultados de los alumnos, 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. Este principio se aplica particularmente a las diferencias de género, ya que existen varias creencias ampliamente extendidas sobre el tipo de preguntas que benefician a los chicos o a las chicas que pueden ser verdad o no para los alumnos que realizan los exámenes del IB. Siempre se deben utilizar pruebas extraídas de evaluaciones anteriores a la hora de emitir este tipo de juicio.

## 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 los alumnos con fines de selección. Si dos alumnos han realizado el mismo examen al mismo tiempo, se puede afirmar con certeza que un alumno con una calificación de 7 ha tenido mejor desempeño ese día que un alumno 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 alumnos obtuvieron una calificación de 6 en Historia, pero respondieron a preguntas distintas o eligieron opciones distintas.
- Dos alumnos, uno de los cuales obtuvo una calificación de 5 en Literatura española en mayo de 2014 y otro que alcanzó una calificación de 5 en la misma asignatura en noviembre de 2012.
- Dos alumnos, uno que obtuvo una calificación de 4 en Física y el otro una calificación de 4 en Química.
- Dos alumnos, uno con una calificación de 4 en Matemáticas y otro con una calificación de 4 en Geografía.
- Un alumno europeo con una calificación de 3 en Informática y una de 6 en Literatura china, y un alumno africano con una calificación de 5 en Literatura japonesa y una de 4 en Biología.
- Dos alumnos de 15 años, uno que obtuvo un certificado del PAI y otro que obtuvo una titulación de otra organización educativa.
- Dos alumnos, uno de los cuales obtuvo una calificación de 6 en Indonesio B NM y otro que alcanzó una calificación de 5 en Indonesio B 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 se evalúan 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. El resultado que obtenga un alumno en la asignatura de Música es un mejor indicador para conocer su predisposición a convertirse en músico profesional que la misma calificación en la asignatura de Artes Visuales, pero ambos resultados pueden ser equivalentes si lo que se quiere predecir es la preparación del alumno para estudiar Historia.

El tema de la equiparabilidad se complica aún más, ya que cada alumno 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 alumnos 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:

- 1. La calidad de trabajo necesaria para lograr las distintas calificaciones en una asignatura o disciplina debe ser equiparable 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 certificación del programa (diploma del IB, certificado del PAI, etc.) sean equiparables.



Aunque el IB procura centrarse en las habilidades de orden superior, las evaluaciones del IB deben ser, en general, equiparables con exámenes similares que ofrecen los países u otros organismos de evaluación.

#### 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) dividen los métodos que existen para comparar las dificultades en dos grandes categorías: los métodos estadísticos y los métodos estimativos.

Los métodos estadísticos se centran en comparar el desempeño de los alumnos en sus 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 de alumnos lo suficientemente grande, el promedio de los resultados debería ser el mismo en ambas evaluaciones.

Por el contrario, los métodos estimativos utilizan a expertos en la materia para estudiar las 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. identifican seis críticas amplias para cada técnica:

#### Críticas a los métodos estadísticos

#### Miden 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 alumnos con algún sesgo inherente?
- Diferencias entre subgrupos: si distintos subgrupos de alumnos se enfrentan a distintos niveles de dificultad, ¿esto no pone en entredicho las conclusiones de la equiparabilidad relativa?
- Desacuerdo entre los métodos: ¿es posible que uno de ellos sea el "correcto"?
- Problemas de forzar la igualdad: ¿cómo afectaría a los alumnos que se enfrentan a la evaluación?

#### Críticas a los métodos estimativos

- Amplitud de criterios: para poder aplicarlos en distintas asignaturas, los criterios deben ser muy amplios, y eso provoca que sean imprecisos.
- Valoración de las respuestas en distintos niveles de exigencia: los examinadores tienden 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: los examinadores tienen 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 de los alumnos típicos.
- Interpretación y contexto: se compara un único examen final con una serie de evaluaciones modulares.
- Suma de varios juicios: la mayoría de las evaluaciones miden varios criterios que se deben equilibrar.

Los partidarios de cada una de estas metodologías suelen ser bastante críticos con la otra. Además, por otro lado, hay quienes argumentan que los enfoques actuales tienen defectos de base. A fin de investigar este tema, Coe et al. aplicaron cinco maneras distintas de medir la equiparabilidad entre asignaturas de dos titulaciones de Inglaterra: el GCSE (General Certificate of Secondary Education) y el GCE A Level (General Certificate of Education Advanced Level). Los autores llegaron a la conclusión de que hay un nivel razonablemente elevado de coherencia entre las maneras 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 la valoración del examinador, el análisis estadístico y los comentarios del profesor (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 (pares de asignaturas, simultaneidad de logros) y de juicios de expertos.

## 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 solo una respuesta correcta: 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 concepto de *pertinencia del constructo* engloba el primero de estos aspectos, ya 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. La otra cara de esta moneda es la autenticidad, que significa que las evaluaciones del IB las componen tareas significativas que presentan situaciones que los alumnos pueden 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 tan auténticas y significativas requieren un alto grado de subjetividad en la corrección, lo que significa que hay que aceptar mayores variaciones entre los examinadores que si, por ejemplo, se utilizasen evaluaciones con preguntas de opción múltiple. Asimismo, esto también afecta a la manejabilidad de la evaluación. Estos tipos de evaluación son más complicados y requieren más tiempo para su preparación y corrección. Además, estas evaluaciones exigen un mayor compromiso por parte de los alumnos para comprenderlas y poder abordarlas. Esto provoca que aumente su carga de trabajo, ya que, por ejemplo, deben 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 a alumnos jóvenes ávidos de conocimiento, informados, instruidos, solidarios y motivados para tener éxito. El IB espera que sus alumnos sean capaces de contribuir a crear un mundo mejor, en el marco del entendimiento mutuo y el respeto intercultural. Todos los programas del IB están comprometidos con el desarrollo de los alumnos según el perfil de la comunidad de aprendizaje del IB.

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 se debe plantear a nivel de programa y debe incluir las normas que rigen la concesión de la titulación completa 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 se debe pensar en el programa de estudios en conjunto que ha cursado el alumno.



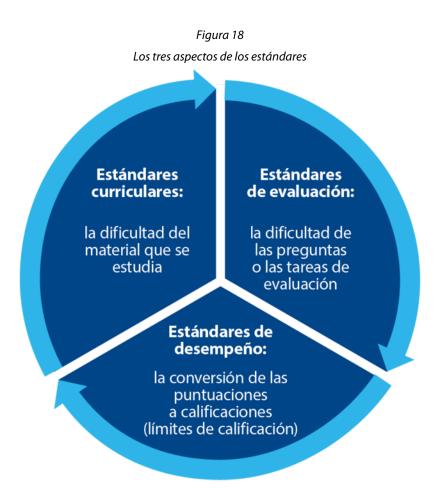
La mayoría de los principios de evaluación que se presentan en este recurso son aplicables a toda la educación del IB, pero en la sección C se aborda cada programa individualmente y se tratan sus peculiaridades.

### Definición de los 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 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 de los alumnos en años anteriores.
- Mantener los estándares resulta tan importante como fijarlos.

## Los tres significados de estándares

Al hablar de *estándares* en el ámbito de la evaluación, normalmente nos referimos a la dificultad de las tareas o de las pruebas que se encomiendan a los alumnos, y son un componente central de la equiparabilidad. Por lo general, el concepto de estándar se puede considerar de tres maneras distintas:



Cada uno de estos tres aspectos puede variar para cambiar así 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 los profesores dispongan del tiempo suficiente para adaptar su enseñanza a los cambios que se produzcan. En el caso del PD y el POP, el plazo ideal serían dos años (el tiempo que dura el curso), pero para el 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 siempre 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 los alumnos, ya que, en cierta medida, estarían familiarizados con esas preguntas la segunda vez. Es posible que en ocasiones se tome la decisión deliberada 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, es posible hacerlo en un examen que ya esté en proceso de elaboración. También en estos casos la equidad exige que se notifique a los profesores 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 el estándar de evaluación. Por ejemplo, si el examen fue más difícil y por ese motivo los alumnos obtuvieron puntuaciones más bajas, se bajarían los límites de calificación para que los alumnos con la misma calidad que los del año anterior obtengan la misma calificación que habrían logrado el año antes. A diferencia de los otros dos tipos de estándares, son sobre todo los examinadores, más que los profesores, quienes deben 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, pero el IB normalmente avisaría con antelación a los colegios para que pudieran gestionar y ajustar lo que se espera de los alumnos. En tal caso, el paso más importante es que los examinadores comprendan 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 alumno demostrar niveles de precisión muy altos. Así, un alumno 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 es el propósito pensado para la evaluación?

La definición de estándar 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 del profesor según lo que quiera que se aprenda. Normalmente se suele prestar más atención a los estándares curriculares y de evaluación con el fin de que las pruebas resulten adecuadas para el alumno y que proporcionen información útil para fundamentar la enseñanza en el futuro.



Figura 19
Es poco probable que de esta evaluación formativa se obtengan comentarios útiles.

## El enfoque normativo y el enfoque basado en criterios en los estándares de desempeño

- El enfoque normativo implica fijar el estándar de desempeño en función de los resultados de los alumnos, por ejemplo, el desempeño del 20 % superior de los alumnos.
- El enfoque basado en criterios implica fijar el estándar de desempeño en función de una descripción de lo que se busca en el desempeño de los alumnos.
- El IB utiliza un 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 de los alumnos en años anteriores.

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

#### ¿En qué consiste el enfoque normativo?

La definición técnica de enfoque normativo se suele relacionar con pruebas estandarizadas. El principio consiste en aplicar una prueba piloto a una muestra típica de alumnos, y utilizar los resultados (a los que, por definición, debe corresponder una distribución normal o curva campaniforme) como escala de referencia para la calificación de cualquier alumno 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 definición 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 alumnos 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 los alumnos por orden según su desempeño y a establecer la proporción de alumnos que recibirá cada calificación. Por ejemplo, los alumnos incluidos en el mejor 15 % de los resultados recibirán la calificación más alta.

#### ¿En qué consiste el 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 establecimiento de estándares de desempeño, ya que dirigió el énfasis hacia medir los logros del alumno "con respecto a un ámbito de conducta bien definido" (Popham, 1978)).

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

Este enfoque presenta limitaciones y dificultades, ya que resulta muy complicado crear descripciones inequívocas y que signifiquen lo mismo para todos los expertos. De hecho, se ha argumentado que "ningún criterio, independientemente de lo bien formulado que esté, permite una interpretación inequívoca" (Wiliam, 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 experto no tiene en cuenta con precisión la exigencia de las preguntas.

#### ¿Qué enfoque utiliza el IB?

El IB utiliza un enfoque flexible basado en criterios, el cual se basa en criterios y también reconoce el efecto Good y Cresswell. En este enfoque, se pide a examinadores expertos que, partiendo de criterios (descriptores de calificaciones finales), establezcan un rango estrecho sobre el que 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 alumno 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
- 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. Esto significa que hay que asegurarse de aplicar los mismos estándares cada año. Es importante tener en cuenta que los estándares del IB se basan en el significado de las calificaciones y no de las puntuaciones. La diferencia entre ambos conceptos se explica

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. Las computadoras 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 esta revisión y, en consecuencia, hay que equilibrar los estándares de evaluación y los de desempeño para que los alumnos no queden 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 cuestionarios de examen 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 cuestionarios de examen 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 los alumnos hagan sus exámenes. Por tanto, se confía en el juicio profesional y la experiencia de los preparadores de exámenes y de los verificadores para crear exámenes coherentes y equilibrados. Posteriormente, los estándares de desempeño se ajustan en función de las respuestas de los alumnos 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.



## La descripción del éxito: logros del alumno en la evaluación sumativa

- La declaración de principios del IB establece el objetivo de formar jóvenes capaces de crear un mundo mejor, por lo cual una buena evaluación deberá respaldar este objetivo.
- Si bien las calificaciones representan una visión muy simplificada de los logros de los alumnos, 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.
- En sus evaluaciones, el IB considera que el juicio profesional es importante para lograr resultados significativos para el alumno, pero también reconoce que ese juicio se debe respaldar con pruebas objetivas con el fin de minimizar los sesgos.
- Los resultados deben ser suficientemente precisos como para no poner en desventaja a los alumnos al ofrecer un resultado que no refleje razonablemente sus logros. Sin embargo, la prioridad es asegurarse de que las evaluaciones resulten significativas.

#### La tiranía de las calificaciones: un mal menor

Pensemos un momento en los conocimientos, las habilidades y la experiencia que se necesitan para convertirse en un chef excelente. El conocimiento serían los ingredientes y las combinaciones de sabores que darán lugar a un plato perfectamente equilibrado. Las habilidades corresponderían a la selección y la preparación de los ingredientes, decidir si se cocina en la parrilla o en el horno, y tomar decisiones sobre la presentación del plato. Finalmente, la experiencia tiene que ver con utilizar las técnicas para alcanzar la perfección, con el juicio de saber cuándo algo está estupendamente cocinado, y con cómo presentar una combinación de ingredientes y platos que produzcan una gran satisfacción en la mesa.



Figura 20
Conocimientos, habilidades y experiencia

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. ¿Serviría de ayuda aumentar la escala de calificación hasta 100, o incluso hasta 1.000? Seguramente la respuesta es que no sirve de mucho. Esto permitiría una mayor variedad de calificaciones para diferenciar entre unos y otros, pero el problema fundamental de intentar comparar conjuntos de habilidades distintos sique presente.

Eso es exactamente lo que pasa en la evaluación: ¿cómo se puede sintetizar en un único resultado la complejidad del conocimiento, la comprensión y las capacidades del alumno? Incluso aunque se pudiera obtener con precisión toda la información sobre el alumno, esto no significa que sea posible concederle una calificación que refleje perfectamente sus talentos.

Otra sugerencia que se suele proponer es no conceder ninguna calificación, sino darle a cada alumno 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 una buena enseñanza y aprendizaje, pero tiene la gran desventaja (aunque algunos de los partidarios de este enfoque lo ven como un punto fuerte) de que es muy difícil hacer comparaciones entre alumnos.

Para intentar resolver este problema, es útil regresar al propósito de la evaluación del IB, que es apoyar a los alumnos para que puedan 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 a los alumnos (normalmente universidades en el caso del PD o el POP, y colegios en el caso del PAI).

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.

Esta cita aparece en la introducción de esta guía, pero vale la pena repetirla. Si va a haber un proceso de selección, el IB tiene la responsabilidad de apoyar a sus alumnos para que las decisiones de selección sean lo más justas y tengan el máximo sentido posible. Si las instituciones receptoras solo reciben informes descriptivos de los alumnos, entonces se encargarán de buscar algún modo de comparar a los alumnos, y casi con total seguridad los sistemas que utilicen serán menos fiables y equiparables que los resultados de los exámenes expresados en calificaciones. Pensemos, por ejemplo, en la validez de una breve entrevista con un tutor y todos los factores que podrían influir en el resultado, sobre los cuales no debería basarse la selección.



Esto no significa que las evaluaciones sumativas sean perfectas, y ni siquiera tal vez una manera especialmente buena de seleccionar a los alumnos, pero sí son una forma más justa que las alternativas. Aún más importante, el IB vela constantemente para que este método sea lo más justo, significativo y fiable posible.

Figura 21
¿Deberíamos hacer distinciones entre los dos alumnos de cada pareja?

El examen indica gue tengo una que tengo una



En el ejemplo precedente, la diferencia entre la primera pareja de alumnos se obtiene 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, seguramente la decisión sería justa, y es probable que, si volvieran a realizar los exámenes, obtendrían los mismos resultados. En la segunda pareja, probablemente la diferencia no resultará significativa, y es probable que ambos alumnos 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 alumnos. Es cierto que para un alumno 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 de los alumnos, la diferencia de calificaciones indica una diferencia significativa en el desempeño.

Esto nos lleva al debate de cuántas calificaciones habría que tener. Si solo existen dos calificaciones (aprobado o no aprobado), la mayoría de los alumnos obtendrá una calificación justa, pero los que estén rozando el límite entre ambas sufrirán consecuencias muy graves. Por el contrario, si hubiera 20 calificaciones distintas, habría muchos más alumnos en un límite de calificación, si bien las consecuencias en este caso serían menos graves. 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 para encontrar un equilibrio justo entre el número de alumnos 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 de los alumnos 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, se debe dar una orientación muy precisa sobre cómo corregir el trabajo de todos los alumnos de manera coherente.

Esto implica hacer un gran hincapié en el juicio profesional de quienes corrigen y, sobre todo, en la pericia profesional de los examinadores supervisores, que son quienes fijan y explican los estándares de corrección. 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

- La corrección es el proceso de evaluar si el alumno ha completado bien la tarea que se le había asignado.
- Popularmente, se considera que corregir consiste en determinar si una respuesta es "correcta" y en otorgarle una puntuación en consecuencia, pero este no es el único enfoque posible. La corrección se puede centrar en los detalles individuales de cada respuesta o adoptar una posición holística.
- Existen otros enfoques de la corrección, el más destacado de los cuales es la valoración comparativa.
- En la evaluación formativa, la corrección puede no conllevar una puntuación numérica, sino ser puramente descriptiva.

## ¿Qué se entiende por corrección?

Figura 22 Por qué la puntuación y el desempeño son cosas distintas.



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

la siguiente imagen, la idea según la cual un camello es un caballo diseñado por un comité es un buen ejemplo de cómo unos criterios (o un esquema de calificación) inadecuados pueden producir malos resultados.

## Figura 23 Es posible que haya que ajustar la respuesta perfecta después de revisar los trabajos de los alumnos.

Pregunta: Diseñe un caballo.



Examinador: "Creo que necesitamos un nuevo esquema de calificación".

La corrección se puede realizar de varias maneras en función de la naturaleza de la tarea. A veces la corrección es muy objetiva: la respuesta del alumno es correcta o no lo es. Esto suele pasar cuando la pregunta solo requiere un par de palabras clave o que el alumno seleccione una respuesta entre varias posibles.

Sin embargo, en otras ocasiones, se trata de algo mucho más subjetivo que requiere que el corrector juzgue si la respuesta del alumno 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. A continuación se muestran parte de las bandas de puntuación del curso de Gestión Empresarial del PD y del curso de Educación Física y para la Salud del PAI.

#### Gestión Empresarial; criterio B: Aplicación

Este criterio evalúa la medida en que el alumno es capaz de aplicar las herramientas, técnicas y teorías de gestión empresarial pertinentes a la organización del estudio de caso.

Puntuación	Descriptor de nivel
0	El trabajo no alcanza ninguno de los niveles especificados por los descriptores que figuran a continuación.
1	Las herramientas, técnicas y teorías de gestión empresarial pertinentes se conectan con la organización del estudio de caso, pero la conexión es inapropiada o superficial.
2	Las herramientas, técnicas y teorías de gestión empresarial pertinentes se conectan apropiadamente con la organización del estudio de caso, pero la conexión no ha sido desarrollada.
3	En general se aplican bien las herramientas, técnicas y teorías de gestión empresarial pertinentes para explicar la situación y las cuestiones de la organización del estudio de caso, aunque la explicación pueda carecer de cierta profundidad o amplitud. Se dan ejemplos.



Puntuación	Descriptor de nivel
4	Se aplican bien las herramientas, técnicas y teorías de gestión empresarial pertinentes para explicar la situación y las cuestiones de la organización del estudio de caso. Los ejemplos son apropiados e ilustrativos.

#### Educación Física y para la Salud (PAI, quinto año); criterio A: Conocimiento y comprensión

Figura 24 Ejemplos de corrección subjetiva en el PD y el PAI

Nivel de logro	Descriptor de nivel
0	El alumno <b>no</b> alcanza ninguno de los niveles especificados por los descriptores que figuran a continuación.
1-2	<ul> <li>El alumno:</li> <li>Indica conocimientos fácticos, procedimentales y conceptuales de Educación Física y para la Salud</li> </ul>
	<ul> <li>Aplica conocimientos de Educación Física y para la Salud para investigar cuestiones y sugerir soluciones a problemas en situaciones conocidas</li> </ul>
	<ul> <li>Aplica terminología relacionada con la actividad física y la salud para transmitir su comprensión de modo poco eficaz</li> </ul>
3-4	<ul> <li>El alumno:</li> <li>Esboza conocimientos fácticos, procedimentales y conceptuales de Educación Física y para la Salud</li> </ul>
	<ul> <li>Aplica conocimientos de Educación Física y para la Salud para analizar cuestiones y resolver problemas en situaciones conocidas</li> </ul>
	<ul> <li>Aplica terminología relacionada con la actividad física y la salud para transmitir su comprensión</li> </ul>
5-6	El alumno:
	<ul> <li>Identifica conocimientos fácticos, procedimentales y conceptuales de Educación Física y para la Salud</li> </ul>
	<ul> <li>Aplica conocimientos de Educación Física y para la Salud para analizar cuestiones y resolver problemas en situaciones tanto conocidas como desconocidas</li> </ul>
	<ul> <li>Aplica terminología relacionada con la actividad física y la salud sistemáticamente para transmitir su comprensión</li> </ul>
7-8	El alumno:
	Explica conocimientos fácticos, procedimentales y conceptuales de Educación Física y para la Salud
	<ul> <li>Aplica conocimientos de Educación Física y para la Salud para analizar cuestiones complejas y resolver problemas complejos en situaciones tanto conocidas como desconocidas</li> </ul>
	<ul> <li>Aplica terminología relacionada con la actividad física y la salud sistemática y eficazmente para transmitir su comprensión</li> </ul>

Otra variación es que la corrección se lleve 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: (1) calidad de la gramática, (2) exactitud de los datos clave, (3) estructura del ensayo y (4) calidad de la conclusión. Para que este enfoque sea satisfactorio, hay que asegurarse de que los criterios sean independientes los unos de los otros. En el ejemplo anterior, si el alumno no hubiera aportado ninguna conclusión, recibiría una penalización en los criterios 3 y 4.

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

# Corrección holística y corrección por criterios

Estos dos enfoques presentan ventajas e inconvenientes. En la impresión global holística suele ser difícil que dos examinadores otorguen una misma puntuación, ya que tienen que sopesar los distintos aspectos del trabajo. Sin embargo, la lista clasificatoria (es decir, la comparación de los trabajos de los alumnos) que se obtiene suele reflejar de manera justa cuál es el mejor trabajo.

Por el contrario, la corrección basada en criterios permite resultados más coherentes, aunque las pequeñas diferencias se pueden magnificar fácilmente. Por ejemplo, si hay tres criterios, cada uno con una puntuación máxima de 4, y un alumno está entre 2 y 3 puntos en cada criterio, dos examinadores razonables podrían dar a ese alumno 6 puntos (2 en cada criterio), 9 puntos (3 en cada criterio) o cualquier puntuación intermedia. Esta es una situación relativamente grave, ya que la puntuación total máxima es de 12.

El otro desafío que presenta la corrección por criterios es que a veces puede originar resultados correctos que no parecen justos. Un ejemplo de esto sería que un alumno haga un trabajo que responde muy bien a lo que se pide en los criterios, y que otro alumno haga un trabajo que es bueno pero que no responde bien a los criterios.

### Otras formas de corrección

La mayoría de la gente se imagina que corregir un trabajo se parece a la imagen que aparece a continuación.

Figura 25
Percepción típica de la corrección



Sin embargo, esta no es la única forma de abordar la corrección. Un enfoque que se ha investigado mucho en los últimos años es la valoración comparativa.

La valoración comparativa se basa en que a la mente humana se le da mejor comparar dos objetos que emitir juicios partiendo de una escala abstracta. Tomemos como ejemplo la primera imagen que aparece a continuación. En una escala del 1 al 10 de temperatura (en la que el 10 es extremadamente caliente), ¿qué puntuación le daría? Ahora compárelo con la segunda imagen. ¿Cuál de estas dos imágenes muestra el elemento más caliente?





Figura 26 ¿Qué es más caliente: un café o un volcán?

Este es un ejemplo muy simple, pero el mismo argumento es válido para todas las decisiones subjetivas, como determinar en qué medida un alumno ha contestado bien a la pregunta del examen en su respuesta, en lugar de comparar cuál de dos respuestas ha contestado mejor a la pregunta.

El enfoque que hay tras la valoración comparativa es que los examinadores toman muchas decisiones sobre qué es comparativamente mejor o peor (contraposición que a menudo se expresa con los términos ganar y perder), y que luego esas decisiones se combinan para crear una clasificación en la que se ordenan los elementos que se evalúan. Las matemáticas no solamente permiten abordar casos inesperados (A gana a B, B gana a C, pero entonces inesperadamente un juez decide que C gana a A), sino también establecer la distancia que debería haber entre dos elementos de una lista ordenada, para lo cual se calcula la probabilidad de cada posible resultado (ganar/perder). Este método se puede utilizar para hallar las puntuaciones tradicionales correspondientes.

Cuando hay varios jueces, el aspecto más importante en la valoración comparativa es que todos deben estar de acuerdo sobre en qué consiste una "buena" respuesta. En muchas situaciones, como en el ejemplo anterior del calor, esto resulta obvio, pero cuando se utiliza para corregir el trabajo de los alumnos es importante ser explícito. Por ejemplo, se considera que un ensayo de Historia es bueno si presenta un argumento convincente respaldado por información precisa, y no si es muy largo y contiene numerosos datos que no están relacionados entre sí. La descripción de en qué consiste algo "bueno" se conoce como declaración de importancia.

El segundo aspecto a tener en cuenta de la valoración comparativa es que las puntuaciones finales no se basan en la valoración de un solo examinador, sino en las valoraciones de cada uno de los examinadores que han corregido el ensayo como parte del proceso. Esto significa que las puntuaciones suelen proceder del consenso de las perspectivas de todos los examinadores, y no de la de uno solo. Este enfoque es muy distinto al habitual que se utiliza en la evaluación y requiere una nueva interpretación del significado de fiabilidad.

La desventaja principal de este método es que requiere muchas más decisiones de corrección. En lugar de considerarse una sola vez, cada elemento pasa por varios análisis, aunque lo habitual es que cada uno de dichos análisis se realice más rápidamente. Por tanto, corregir mediante la valoración comparativa requiere más tiempo, y para resolver este problema se ha desarrollado una variante denominada valoración comparativa adaptativa. Dicha variante procura obtener un resultado coherente a partir de menos decisiones del tipo "ganar o perder", para lo cual se centra en las valoraciones que son verdaderamente necesarias.

Es probable que la valoración comparativa aporte beneficios al IB en las tareas más auténticas y significativas que son complicadas de corregir de manera fiable. Para obtener más información sobre la valoración comparativa y la valoración comparativa adaptativa, consulte la "Bibliografía".

# Corrección y 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 los parecidos y las diferencias entre ambos. Esto dificulta mucho comparar dos respuestas para determinar cuál es la mejor, pero en el caso de la evaluación formativa, cuyo objetivo es aportar comentarios para respaldar el aprendizaje, es posible que esto no sea necesario.



# ¿Qué es una buena evaluación?

- No existe una única respuesta a esta pregunta. Todo depende de la importancia relativa que se dé a las distintas prioridades, y del propósito de la evaluación.
- 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 la enseñanza y el aprendizaje.
- Por regla general, las evaluaciones deben incluir una variedad de tareas y permitir organizar actividades de mayor profundidad en clase, además de exámenes.
- Existen muchos detalles técnicos sobre qué significa la validez de una evaluación, pero básicamente se trata de ponderar bien los objetivos de la evaluación.

En realidad, esta pregunta tan sencilla resulta muy difícil de responder. El motivo es que cada persona tiene distintas prioridades, por lo que no hay respuestas correctas o incorrectas.

Para el IB, el principio subyacente 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 el resto de consideraciones, como la fiabilidad y la carga de trabajo del alumno.

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

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

A continuación se abordan uno por uno estos aspectos.

# Una buena evaluación sirve de apoyo a los objetivos curriculares

- La evaluación debe fomentar la buena enseñanza (repercusión positiva).
- La evaluación debe ser previsible.

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

El IB parte del principio de que el diseño de las evaluaciones debe fomentar los resultados educativos más deseados para los alumnos. Los efectos sobre el aprendizaje de los alumnos continúan siendo una consideración esencial en el diseño de las evaluaciones y, junto con la pertinencia del constructo (es decir, evaluar lo correcto), son la prioridad del IB para decidir cómo encontrar el equilibrio entre los distintos elementos de la validez.

La evaluación de alto impacto tiene una considerable influencia sobre la enseñanza y el aprendizaje. Para aprovechar esto, se pueden diseñar instrumentos de evaluación que fomenten una buena pedagogía y la

implicación constructiva de los alumnos en su propio aprendizaje, teniendo en cuenta a la vez las corrientes de pensamiento más recientes sobre teoría del aprendizaje (por ejemplo, Murphy, 1999).

Las características personales que se desea que tengan los alumnos, expresadas en la declaración de principios del IB, se ajustan muy bien a la teoría constructivista del aprendizaje, según la cual los alumnos se implican activamente en el proceso de aprendizaje, se responsabilizan de su propio aprendizaje y amplían sus conocimientos, comprensión y destrezas a través de la investigación.

En los requisitos de evaluación de varias asignaturas, se espera que el alumno muestre empatía respecto a perspectivas culturales distintas de la suya propia. Las cualidades de naturaleza afectiva, como son la solidaridad y la compasión, resultan más difíciles de incluir en la evaluación formal, pero deben, no obstante, estar representadas en el sistema de evaluación global. Esto se consigue parcialmente mediante elementos curriculares que no se evalúan, como el elemento de servicio comunitario del PAI o el requisito de Creatividad, Actividad y Servicio (CAS) del PD y el POP. Asimismo, las evaluaciones del IB muestran prácticas de trabajo éticas y una comprensión y apreciación de las diferencias.

En lo que se refiere al diseño de los cursos, 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 de los alumnos.

Aunque el IB es muy consciente de que las evaluaciones tienen un efecto de repercusión en la enseñanza y el aprendizaje, también insta a los colegios a adoptar pedagogías que desarrollen en los alumnos 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. Si quiere obtener más información sobre los estudios que realiza el IB, consulte la sección "Investigación" del sitio web.

#### ¿Qué es la buena predictibilidad?

El término predictibilidad designa la capacidad de estimar qué sucederá o cuándo sucederá algo. En el ámbito de la evaluación, esto se refiere a la capacidad de los colegios para determinar qué preguntas aparecerán en los exámenes y cuándo. Es fundamental que las prácticas del IB en materia de evaluación tengan una buena predictibilidad, ya que, al cumplir este principio, el IB muestra que respeta los requisitos de sus constructos tal como se publican para los profesores, lo cual da pie a una oportunidad de evaluación "justa" en lo referido al cumplimiento curricular. Lo que el IB haya dicho que se va a evaluar, se evaluará.

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. Se presta atención para eliminar el riesgo de mala predictibilidad hacia el final de un curso, 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 debería ser una sorpresa, ni para el alumno ni para el colegio. 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, el IB procura eliminar los problemas causados por la predictibilidad. Para ello, las competencias se miden mediante evaluaciones diseñadas específicamente para no dar ventajas a aquellos alumnos que lleguen con respuestas preparadas con antelación (por ejemplo, se busca un equilibrio entre el conocimiento, la comprensión y el uso de un estímulo no visto anteriormente).

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.

Cuando el modelo de evaluación lo permite, es aceptable que los temas que estén levemente relacionados se evalúen en distintas pruebas. Esto es coherente con la educación del IB, que está en contra de aislar las asignaturas y las opciones.



Figura 27 ¿Qué es la buena y la mala predictibilidad?

Buena predictibilidad	Mala predictibilidad
Si se hiciese 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 que formulase el IB para esa asignatura.	Los profesores preparan en exceso a sus alumnos 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. Los instrumentos o componentes de evaluación pueden adoptar la forma de una prueba de examen, una carpeta, un proyecto o un trabajo de investigación. Los conceptos de componente y de tarea de evaluación se solapan. A veces, un alumno puede realizar para un componente una sola tarea de entre varias a elegir.

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 a cuyas instituciones la mayoría de los alumnos del IB pretendían acceder", y este principio se extiende al POP y al PAI. 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 eguidad en la evaluación (Linn, 1992; 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 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 a los alumnos en áreas que no resultan adecuadas para los exámenes. Su principal ventaja es que permiten pedir a los alumnos 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 y las investigaciones matemáticas. Los exámenes orales, en

los cuales un profesor tiene que plantear preguntas y reaccionar a las palabras del alumno, también requieren que se hagan en clase, si bien esto puede cambiar en el futuro debido al desarrollo de la tecnología informática.

El trabajo evaluado internamente tiene otras ventajas en el marco de las titulaciones internacionales. Dicho trabajo permite mucha flexibilidad en la elección de los temas y, a la vez, seguir abordando un conjunto de destrezas comunes. Los colegios pueden así situar el estudio en un contexto cultural o geográfico local, o establecer lazos más estrechos entre el aula y el mundo exterior que la rodea. Los colegios internacionales, cuyos alumnos a menudo proceden 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. Por otra parte, la evaluación interna puede utilizarse de distintos modos para desarrollar vínculos con culturas lejanas, generalmente comunicándose con colegios de otras partes del mundo. Brown (2002) también señala el valor de la evaluación interna en el Programa del Diploma a la hora de dar cabida a la diversidad cultural, lo que fomenta "una perspectiva más amplia del internacionalismo", permitiendo una multiplicidad de enfoques culturales y ofreciendo a cada alumno la oportunidad de experimentar diversos valores culturales.

Además, la evaluación interna a menudo ofrece a los alumnos la oportunidad de elegir sus propios temas o cuestiones, según sus intereses particulares, y les permite mayor control de su propio aprendizaje. Esta flexibilidad de enfoque hace que la evaluación interna sea una valiosa adición a la educación de los alumnos, 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. Una de las dos principales dificultades es que resulta mucho más difícil garantizar que los alumnos no cometan conductas improcedentes (por ejemplo, que otra persona les haga el trabajo). La disponibilidad de Internet agrava aún más la situación. El colegio se encuentra en la mejor posición para identificar los casos en los que el alumno no hace el trabajo que entrega. Como ayuda para a resolver este problema, el equipo directivo del colegio tiene a su disposición varios recursos en el sitio web del IB.

La segunda dificultad principal es que la evaluación realizada en clase puede suponer una carga de trabajo importante tanto para el profesor como para el alumno. 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 alumno les dedique mucho tiempo. Aunque es correcto que los profesores empleen mucho tiempo preparando a los alumnos en las destrezas y procesos necesarios para la evaluación interna, puede existir una fuerte tentación, por parte del alumno y el profesor, de ensayar y practicar la tarea concreta establecida para la evaluación interna más de lo necesario, a fin de hacerlo lo mejor posible. Esto implica reducir aún más el tiempo dedicado a la enseñanza.

La evaluación realizada en clase se puede corregir internamente o externamente. En el caso de la evaluación externa, el trabajo que hace el alumno se envía al IB y lo corrige un examinador. La calidad de las decisiones del examinador se puede supervisar como se hace habitualmente (véase la sección "Modelo de calidad"). La evaluación realizada en clase puede conllevar que haya trabajos extensos difíciles de corregir.

En el caso de la evaluación interna, el colegio (normalmente el profesor del alumno) corrige el trabajo del alumno. Posteriormente, el IB, mediante un proceso de moderación, comprueba que el profesor haya aplicado correctamente el estándar global.

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

Asimismo, la opinión del profesor puede verse influida si conoce trabajos previos del alumno, ya que esto crea ciertas expectativas. Puede que los profesores a veces no tengan claros los límites de su papel al orientar y apoyar a los alumnos mientras llevan a cabo el trabajo de evaluación interna, y es posible que tengan solo una visión limitada de los estándares de logro que hay a nivel mundial en su asignatura. Cuando evalúan el trabajo de sus propios alumnos, los profesores pueden verse muy influenciados por los



estándares generales existentes en su colegio. Incluso cuando no hay ningún incentivo para conceder a los alumnos puntuaciones más altas de lo que correspondería, la relación profesional que los profesores establecen con ellos provoca que sea difícil tomar decisiones objetivas. Los estudios realizados sobre este sesgo inconsciente indican que puede tener un efecto notable.

Para el IB, la utilización de tareas de evaluación significativas que impliquen necesariamente un trabajo realizado en clase, y que deba corregir alguien que haya sido testigo del desarrollo de los alumnos, conlleva ventajas que superan a los riesgos de la evaluación interna. Por ese motivo, normalmente las tareas de evaluación interna forman parte de cualquier conjunto de evaluaciones.

Por último, las tareas en el aula también pueden dar a los alumnos la oportunidad de demostrar aspectos de la educación del IB que no se evalúan.

#### Trabajo colaborativo y puntuaciones individuales

Una cuestión complicada es cómo evaluar las tareas colaborativas, en las que se espera que los alumnos realicen un trabajo colectivo para completar una actividad. Este es un aspecto importante que los alumnos encontrarán 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 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 estas situaciones, 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 alumno al resultado total. Es posible que un alumno 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 de los alumnos. 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 uno de los alumnos.

Este enfoque puede dar lugar a sesgos culturales, ya que el individualismo es una ideología tradicionalmente propia de la Europa occidental. El IB es consciente de este problema, pero considera que, incluso en otras culturas, los resultados de la evaluación se suelen usar para tomar decisiones de selección individuales.

# Una buena evaluación considera las competencias más amplias del alumno 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 de los alumnos. Este arraigado objetivo del IB se refleja actualmente en iniciativas como las que adoptan diversos gobiernos para dotar a los alumnos con habilidades del siglo XXI y competencias laborales.

La filosofía del IB y su enfoque acerca de las competencias del alumno se centran en el perfil de la comunidad de aprendizaje del IB y su relación con la mentalidad internacional. Esto se explicará más detalladamente en la sección "Elementos comunes a todos los programas".

Una buena evaluación tiene en cuenta toda la variedad de resultados que el curso pretende lograr y permite que los alumnos demuestren 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 de los alumnos, y de su relación con la mentalidad internacional.

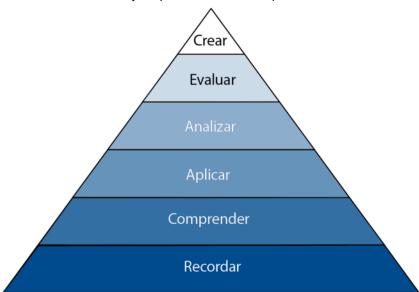
#### Habilidades cognitivas de orden superior

El objetivo de las evaluaciones del IB es 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 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 mundial para directores de colegio del IB celebrada en Singapur en 2011, llevó esta argumentación 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, y que lo que necesitan los ciudadanos del siglo XXI es la capacidad de analizar, interpretar y seleccionar conocimientos (Mitra, 2011).

En las evaluaciones del IB, hace tiempo que se intenta prestar mucha atención a las llamadas habilidades cognitivas "de orden superior" (Bloom *et al.*, 1956; Anderson y Krathwohl, 2001). Puede que exista desacuerdo sobre la naturaleza jerárquica de los niveles que propone Bloom o sobre el número de niveles, pero su taxonomía de objetivos educativos continúa proporcionando un marco útil a través del cual expresar la diversidad de capacidades que se requiere. 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 de los alumnos solo pueden valorarse adecuadamente si se les exige analizar, sintetizar y evaluar con cierta amplitud. La evaluación del desempeño es el único modo realista de medir los logros de los alumnos 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 28
Una posible manera de describir una jerarquía de habilidades de pensamiento es la taxonomía de Bloom.



Para que sean pertinentes, las evaluaciones del IB deben ser capaces de reconocer y recompensar el desempeño de los alumnos 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.

# Las competencias del alumno y 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.

Cada vez se afirma más que este siglo requiere habilidades radicalmente distintas a las que se esperaban en las generaciones anteriores. Si bien hay quienes argumentan que el enfoque de indagación que subyace a las habilidades del siglo XXI se ha valorado desde Sócrates, existe cierto consenso general sobre la importancia de proporcionar a los alumnos una amplia variedad de atributos para prepararlos para la vida. Véanse, por ejemplo, los argumentos que se dan en Llewellyn *et al.*, 2014.

Hay muchas maneras distintas de categorizar estas habilidades, por ejemplo, el marco de competencias para el siglo XXI de la OCDE, el que propone RAND Education, y el del National Research Council (NRC). El IB, por su parte, describe estas competencias dentro del perfil de la comunidad de aprendizaje.

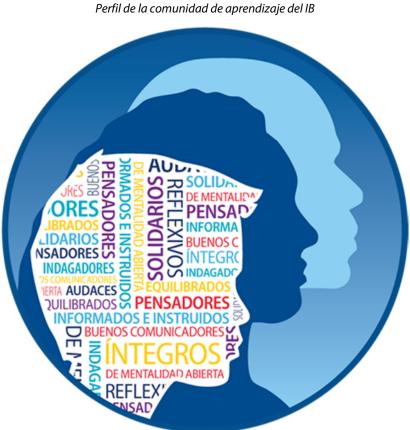


Figura 29 Perfil de la comunidad de aprendizaje del IB

No todos los aspectos del perfil de la comunidad de aprendizaje se pueden medir adecuadamente con la evaluación sumativa, 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, aunque no esté diseñada para medirlas, puede ofrecer a los alumnos 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 e introducir contextos inesperados a los alumnos.

Si quiere obtener más información sobre el enfoque amplio del IB respecto a las competencias de los alumnos, consulte el material disponible en el sitio web del IB o el documento *De los principios a la práctica* del programa pertinente.

#### Mentalidad internacional y entendimiento intercultural

Los alumnos que estudian los programas del IB se encuentran en numerosos países y tienen una amplia variedad de nacionalidades. Además de los objetivos académicos de nuestros programas, el IB pretende que los alumnos se formen como "jóvenes solidarios capaces de contribuir a crear un mundo mejor y más pacífico, en el marco del entendimiento mutuo y el respeto intercultural" y alentarlos a "entender que otras personas, con sus diferencias, también pueden estar en lo cierto" (declaración de principios del IB, 2002). Por lo tanto, en la enseñanza el IB hay un contexto internacional y un propósito de comprensión intercultural, y ambos elementos deben reflejarse en la evaluación.

Lo más importante para lograr esto es emplear a expertos académicos —entre otros, los autores de los cuestionarios de examen y los encargados del desarrollo curricular— procedentes de distintos contextos culturales. Para cumplir los principios del IB es importante no ocultar las diferencias, sino utilizarlas de manera que los alumnos puedan explorarlas sin verse desfavorecidos.

En algunas áreas disciplinarias, para fomentar la variedad cultural se pueden tener en cuenta distintas perspectivas culturales en el currículo. Podemos encontrar ejemplos de este enfoque en Biología, Química, Psicología y Artes Visuales. En las tres primeras, las estructuras de opciones 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, la mentalidad internacional se potencia a través de los materiales y las fuentes de inspiración que se fomenta que use el alumno. Algunos ejemplos en este sentido son las asignaturas de Artes, Literatura y Lengua, pero esta noción también se puede incluir en una amplia variedad 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, ya que esta se centra en los logros más que en los atributos afectivos.

En el IB esto se aborda con los elementos que no se evalúan de sus programas, como Creatividad, Actividad y Servicio (CAS) en el PD, y el Proyecto Comunitario en el PAI. Como los alumnos solo pueden obtener el título del correspondiente programa si completan los elementos no evaluados, estos tienen un impacto significativo en el resultado global de la evaluación del IB.





Tal vez se considere que permitir a los alumnos elegir qué preguntas quieren responder es la mejor manera de abordar los distintos requisitos internacionales en materia de evaluación, pero esto provoca problemas en lo que se refiere a mantener la equiparabilidad entre las opciones. Esto siempre sucede cuando se pueden elegir las preguntas, o cuando las tareas de evaluación son muy abiertas. Resulta complicado siquiera definir en qué consiste una "exigencia igual" cuando los alumnos tienen bagajes educativos muy distintos. En general, la manera más fácil de mantener la equiparabilidad es establecer tareas comunes que permitan a los alumnos apoyarse en sus propias experiencias para dar las respuestas. En esos casos, el examinador afronta el reto de mantener un estándar común, si bien esto resulta un poco más fácil que corregir según el mismo estándar dos tareas que pueden tener niveles de exigencia diferentes.

Para obtener información sobre la equiparabilidad, se puede analizar el desempeño de los alumnos. 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. Las preguntas que se refieren a deportes, viajes, entretenimiento, acontecimientos históricos, incluso al tiempo meteorológico, deben prepararse con sumo cuidado. Puede parecer que el único modo de evitar este problema sea preparar preguntas de examen carentes de todo contexto sociocultural excepto el mínimo común denominador. Sin embargo, hacerlo así no solo limitaría mucho las preguntas de exámenes y las haría aburridas, sino que iría en contra de toda la filosofía de evaluación del IB y contra el correcto ejercicio de la actividad de evaluación en lo concerniente a garantizar su validez mediante tareas basadas en contextos. La evaluación y el trabajo contextualizados son esenciales para el buen aprendizaje.

Existen dos posibles soluciones para este dilema. En primer lugar, puede proporcionarse a los alumnos información contextual complementaria. 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 la misma pregunta de examen (siempre que no ocupe demasiada extensión, con el fin de evitar que distraiga a los alumnos de la finalidad de la evaluación).

Un segundo método consiste en utilizar preguntas y tareas de evaluación más abiertas que permitan a los alumnos elegir el contexto en el que responder. En este enfoque, la puntuación ha de centrarse en niveles de comprensión más profundos, y no en conocimientos directos del contenido de la asignatura, ya que no habrá base común de contenido. Esto está muy en la línea de la filosofía de evaluación del IB.





Figura 31
Variedad de normas y contextos culturales

Incluso aplicando estos dos métodos, los alumnos pueden encontrarse con tareas de evaluación cuyos contextos no les resulten familiares en su propio medio sociocultural. También esto está en la línea de la filosofía de evaluación del IB, dado que uno de los objetivos de los programas es abrir las mentes de los alumnos a otras formas de hacer las cosas, aumentar su conciencia global y su competencia para operar en un entorno cultural que no les resulte familiar. Parte del requisito del pensamiento abstracto es que los alumnos deben poder aplicar sus conocimientos en situaciones desconocidas. Es apropiado incluir tales elementos en la evaluación, siempre que afecten por igual a alumnos de diferentes culturas.

Un número significativo de alumnos del IB se presentan a exámenes en una lengua que no es aquella en la que son más competentes. En casi todos estos casos se trata del inglés, porque los alumnos que estudian en francés o español (las otras dos lenguas en las que se realiza la evaluación) suelen ser hablantes nativos. Debe procurarse de manera especial que la redacción de las preguntas no sitúe en desventaja a los alumnos que realizan la evaluación en su segunda lengua. De esto se encargan los responsables de la edición de exámenes.

La evaluación sumativa del IB, al igual que la gran mayoría de sistemas de evaluación formales, es muy individualista. Como señaló Brown (2002), esto se debe en gran medida a que el Programa del Diploma pertenece a la tradición europea occidental, y las sociedades europeas occidentales son individualistas por naturaleza. Se evalúa a los alumnos casi exclusivamente sobre lo que logran por sí mismos. Puede decirse que culturalmente esto carece de equidad, ya que hay algunas culturas en las que lo que aporta el individuo está siempre subordinado a lo que aporta un grupo más amplio: lo que importa es lo que logra el grupo. También sucede que, en términos de equidad individual, hay algunas personas que trabajan mejor en equipo que individualmente, y viceversa. Además, es práctica habitual que, tanto en el aula como en el mundo laboral, los individuos trabajen de forma interdependiente más que independiente.

## ¿Cómo es una buena evaluación en pantalla?

Anteriormente se ha explicado que la evaluación en pantalla puede ofrecer evaluaciones de mejor calidad gracias al uso de la computadora como medio para plantear preguntas que resultarían imposibles en un examen en papel. Esto nos lleva a tres aspectos claros que definen cómo debe ser una buena evaluación en pantalla:

- La evaluación en pantalla utiliza las oportunidades que ofrecen las computadoras para crear tareas más válidas, ya sea por su autenticidad y pertinencia, porque se reducen los sesgos o debido a cualquier otro factor.
- Elimina las barreras que dificultan a determinados alumnos realizar la evaluación, o mejora el proceso general; por ejemplo, al eliminar los posibles riesgos y retrasos que conlleva el envío de exámenes en papel.
- No crea nuevas barreras para los alumnos que realizan la evaluación, como podrían ser interfaces difíciles de utilizar o problemas técnicos.

Por el contrario, si se utilizan características (como videos) simplemente porque están disponibles y no para mejorar la validez de la evaluación, esto denota que la evaluación en pantalla es de poca calidad. La tecnología puede y debe respaldar el proceso de evaluación, pero no determinar la dirección que se le da.



# Principios de evaluación del IB

Los cinco puntos siguientes resumen los principios subyacentes de la evaluación del IB.

Las evaluaciones del IB deben:

- Resultar válidas para los propósitos para los que están pensadas. Esto significa que debe haber un equilibrio entre las demandas opuestas de la pertinencia del constructo, la fiabilidad, la equidad (es decir, que no haya sesgos), la equiparabilidad con otras evaluaciones, y la manejabilidad para los alumnos, los colegios y el IB.
- Tener un efecto de repercusión positivo: su diseño debe favorecer una enseñanza y un aprendizaje de buena calidad.
- Resultar adecuadas para el mayor número de alumnos posible, de manera que puedan demostrar su nivel de logro personal.
- Formar parte integrante del programa del IB, y no considerarse como elementos aislados. Hay que preguntarse si las evaluaciones favorecen la simultaneidad del aprendizaje y la experiencia del alumno
- Respaldar los principios del IB y las competencias de los alumnos, sobre todo para que sean indagadores, informados e instruidos, pensadores, buenos comunicadores y tengan una mentalidad internacional.

# ¿Qué se entiende por *práctica*?

- Las prácticas de evaluación del IB engloban las evaluaciones sumativas realizadas en el PD, el POP y
  el PAI.
- Estas no se aplican al PEP, en el que el IB 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 de los alumnos que realizan 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 docente para utilizarla en el PEP.

Un principio establece *por qué* se hace algo, mientras que una práctica describe *cómo* se hace. Por eso, 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 que describen los pasos individuales para aplicar cada práctica. Cuando estos procesos conciernen a los colegios, se explican en detalle en los *Procedimientos de evaluación* específicos de cada programa.



# Comunicación de los logros de los alumnos

- La declaración de principios del IB establece el objetivo de formar jóvenes capaces de crear un mundo mejor, por lo cual 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 de los alumnos, 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 de los alumnos no se reducen a los resultados obtenidos en los exámenes, así que, incluso cuando se trate de calificaciones finales, es importante recordar que lo que para un alumno puede suponer un resultado decepcionante, para otro será un gran logro.

El Bachillerato Internacional tiene como meta formar jóvenes solidarios, informados y ávidos 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, 2002.

# ¿Qué significan las calificaciones finales del IB?

Los resultados de un alumno que realiza las evaluaciones del IB se denominan calificaciones finales. Estas calificaciones describen la calidad del trabajo que el alumno ha demostrado en sus respuestas.

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

#### Figura 32

Ejemplos de descriptores de calificaciones finales del Programa del Diploma

Programa del Diploma

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

#### Calificación 7

El alumno demuestra consciencia 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; consciencia 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 consciencia 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 programa, 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 los educadores 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 33 ¿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) (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

$$A = \int_{1}^{e} l dx - \int_{1}^{e} \ln x dx$$

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

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

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

En el contexto de la evaluación del IB, el argumento nos lleva a la validez de los propósitos de las calificaciones finales. Las calificaciones finales están pensadas para permitir que las partes interesadas puedan comparar los logros de los alumnos, así que tiene sentido utilizar métodos estadísticos y cualitativos para intentar lograr la paridad en el significado de las calificaciones.

# ¿Cuál es la diferencia entre puntuaciones y calificaciones?

Puntuación y calificación son conceptos distintos.

Figura 34
No importa solo lo lejos que se llegue caminando, sino también cuál es el terreno.



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

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



Existen muchas metáforas para explicar la diferencia entre las puntuaciones y las calificaciones. Por ejemplo, en la imagen anterior, la distancia caminada se consideraría la puntuación, ya que es una medida habitual de lo lejos que alguien ha llegado. Sin embargo, para entender el logro que eso representa, hay que tener en cuenta el terreno por el que se estaba caminando, y eso es precisamente lo que se hace cuando se otorgan calificaciones.

- Las puntuaciones representan en qué medida el alumno ha completado la tarea.
- Las calificaciones tienen en cuenta la dificultad de la tarea para indicar en qué medida debería impresionarnos la puntuación del alumno.

Fijémonos en los dos ejemplos siguientes. En el primer ejemplo se esperaría que un alumno de 16 años realizase casi toda la tarea correctamente para recibir una "buena" calificación, mientras que en el segundo ejemplo se esperaría bastante menos para que el alumno lograse una "buena" calificación.





Figura 35
Las expectativas respecto a los buenos alumnos son menores en las tareas difíciles que en las fáciles.

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

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

Esto conduce a uno de los desafíos que se plantean al establecer una evaluación de buena calidad. Los alumnos deben 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 la mitad menos brillante de los alumnos no pueda ni siquiera empezar la tarea, por lo cual no se podrán utilizar las calificaciones finales para hacer una distinción entre estos alumnos.

El otro riesgo que se deriva de contar con tareas demasiado sencillas es el de comenzar a medir la precisión en vez de la comprensión. De esa manera, los alumnos que cometieran errores menores no recibirían las calificaciones más altas aunque tuvieran una buena comprensión del tema, y esto no es lo que pretende la evaluación.

#### La tiranía de las calificaciones: un mal menor

Pensemos un momento en los conocimientos, las habilidades y la experiencia que se necesitan para convertirse en un chef excelente. El conocimiento serían los ingredientes y las combinaciones de sabores que darán lugar a un plato perfectamente equilibrado. Las habilidades corresponderían a la selección y la preparación de los ingredientes, decidir si se cocina en la parrilla o en el horno, y tomar decisiones sobre la presentación del plato. Finalmente, la experiencia tiene que ver con utilizar las técnicas para alcanzar la perfección, con el juicio de saber cuándo algo está estupendamente cocinado, y con cómo presentar una combinación de ingredientes y platos que produzcan una gran satisfacción en la mesa.



Figura 20
Conocimientos, habilidades y experiencia

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. ¿Serviría de ayuda aumentar la escala de calificación hasta 100, o incluso hasta 1.000? Seguramente la respuesta es que no sirve de mucho. Esto permitiría una mayor variedad de calificaciones para diferenciar entre unos y otros, pero el problema fundamental de intentar comparar conjuntos de habilidades distintos sique presente.

Eso es exactamente lo que pasa en la evaluación: ¿cómo se puede sintetizar en un único resultado la complejidad del conocimiento, la comprensión y las capacidades del alumno? Incluso aunque se pudiera obtener con precisión toda la información sobre el alumno, esto no significa que sea posible concederle una calificación que refleje perfectamente sus talentos.

Otra sugerencia que se suele proponer es no conceder ninguna calificación, sino darle a cada alumno 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 una buena enseñanza y aprendizaje, pero tiene la gran desventaja (aunque algunos de los partidarios de este enfoque lo ven como un punto fuerte) de que es muy difícil hacer comparaciones entre alumnos.

Para intentar resolver este problema, es útil regresar al propósito de la evaluación del IB, que es apoyar a los alumnos para que puedan 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 a los alumnos (normalmente universidades en el caso del PD o el POP, y colegios en el caso del PAI).

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.

Esta cita aparece en la introducción de esta guía, pero vale la pena repetirla. Si va a haber un proceso de selección, el IB tiene la responsabilidad de apoyar a sus alumnos para que las decisiones de selección sean lo más justas y tengan el máximo sentido posible. Si las instituciones receptoras solo reciben informes descriptivos de los alumnos, entonces se encargarán de buscar algún modo de comparar a los alumnos, y casi con total seguridad los sistemas que utilicen serán menos fiables y equiparables que los resultados de los exámenes expresados en calificaciones. Pensemos, por ejemplo, en la validez de una breve entrevista con un tutor y todos los factores que podrían influir en el resultado, sobre los cuales no debería basarse la selección.



Esto no significa que las evaluaciones sumativas sean perfectas, y ni siquiera tal vez una manera especialmente buena de seleccionar a los alumnos, pero sí son una forma más justa que las alternativas. Aún más importante, el IB vela constantemente para que este método sea lo más justo, significativo y fiable posible.

Figura 21
¿Deberíamos hacer distinciones entre los dos alumnos de cada pareja?

El examen indica

Gue tengo una



En el ejemplo precedente, la diferencia entre la primera pareja de alumnos se obtiene 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, seguramente la decisión sería justa, y es probable que, si volvieran a realizar los exámenes, obtendrían los mismos resultados. En la segunda pareja, probablemente la diferencia no resultará significativa, y es probable que ambos alumnos 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 alumnos. Es cierto que para un alumno 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 de los alumnos, la diferencia de calificaciones indica una diferencia significativa en el desempeño.

Esto nos lleva al debate de cuántas calificaciones habría que tener. Si solo existen dos calificaciones (aprobado o no aprobado), la mayoría de los alumnos obtendrá una calificación justa, pero los que estén rozando el límite entre ambas sufrirán consecuencias muy graves. Por el contrario, si hubiera 20 calificaciones distintas, habría muchos más alumnos en un límite de calificación, si bien las consecuencias en este caso serían menos graves. 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 para encontrar un equilibrio justo entre el número de alumnos que se encuentren en un límite de calificación y las consecuencias que tendría recibir una calificación equivocada.

# Oué se considera una convocatoria de exámenes satisfactoria

Figura 36 Lo que define si una convocatoria de exámenes es satisfactoria o no depende del punto de vista.



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

Pudimos elegir a los alumnos adecuados para nuestros cursos.



colegios cercanos.

A nuestros alumnos les ha No hubo errores en las ido mejor que a los de los preguntas ni en la corrección.

Determinar si algo es satisfactorio depende del punto de vista de cada uno, y todos los puntos de vista son igual de importantes. El IB hace hincapié tanto en la validez general como en la facilidad de uso:

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

# Los logros no se limitan a las calificaciones

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?.

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 solidarios o de audaces (suponiendo que sea razonable darles un valor numérico a estas cualidades esenciales)? Sin embargo, 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 37 Los logros de un alumno van mucho más allá de lo que se puede demostrar.

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.

En las evaluaciones, el IB solo registra los logros finales del alumno, sin precisar lo difícil que le resultó al alumno 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 del

Si bien existen métodos que permiten calcular el "valor añadido" de los logros del alumno o "calificaciones previstas basadas en los logros anteriores", el IB recomienda proceder con cautela cuando se utilicen para hacerse una idea del éxito de un alumno. Estas mediciones se basan en un alumno promedio, pero cabe recordar que cada alumno 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 durante un programa completo.

# 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 expertos 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 38 Responsabilidades de los actores principales del ciclo de evaluación

	_	_	_		
Exar	nina	dor	prin	cipal	

- Responsable de un componente.
- Toma la decisión final sobre qué puntuación deben recibir las respuestas de los alumnos en ese componente.
- Se asegura de que todos los examinadores comprendan el estándar de corrección en ese componente.
- Guía al examinador 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 de los alumnos siguiendo el estándar que fija el examinador principal.

#### ΙB

- Responsable de todos los aspectos de la evaluación.
- Responsable de los procesos de evaluación, entre los que se incluyen la contratación de examinadores, el control de calidad del trabajo de los examinadores y la publicación de resultados.
- Toma decisiones sobre cuestiones de conducta improcedente, la mala administración, o disposiciones o consideraciones especiales.
- Acepta o rechaza los límites de calificación que recomienda el examinador jefe.

# Examinador principal y examinador jefe

El examinador principal se encarga de un componente (por ejemplo, la evaluación interna o la prueba 2) y sus principales responsabilidades son:

• Decidir a qué respuestas se les conceden puntos (fijar el estándar de corrección).



- Explicar dicho estándar a su equipo de examinadores, y revisar el esquema de calificación y cualquier otra orientación para los examinadores, incluidos los exámenes de práctica.
- Guiar al examinador jefe para establecer los límites de calificación para su componente.

En resumen, el examinador principal es la persona encargada de las cuestiones académicas que giran en torno a la corrección de un examen concreto. El examinador principal cuenta con el apoyo de varios examinadores con experiencia (los examinadores supervisores), que discuten las cuestiones y le aportan recomendaciones.

La función del examinador principal no conlleva necesariamente formar parte del equipo de autores del examen, si bien esto sería muy poco común.

El examinador principal trabaja para el IB durante la convocatoria de exámenes, pero no forma parte del personal de la organización. Para procurar una mayor coherencia, normalmente se cuenta con el mismo examinador principal durante varias convocatorias de exámenes.

El examinador jefe es el responsable de mantener la calidad de varios componentes relacionados entre sí. En el PD y el POP esto significa que es el responsable de una asignatura completa, mientras que en el PAI (que solo tiene un componente por disciplina) el examinador jefe es el responsable de un grupo de asignaturas, como Ciencias o Individuos y Sociedades. El examinador jefe actúa como experto académico del IB en su área y soluciona cualquier diferencia que haya entre los examinadores principales.

El examinador jefe es responsable de los siguientes aspectos:

- Garantizar que los estándares resulten adecuados dentro de los componentes y entre los componentes.
- Asegurar la coherencia entre los componentes, tanto en la corrección como en la redacción de los exámenes.
- Solucionar cualquier desacuerdo que pueda haber entre los examinadores principales o con ellos.
- Dirigir el proceso de concesión de calificaciones finales y recomendar al IB los límites de calificación finales.
- Actuar como embajador en nombre del IB.

Además, se invita a los examinadores jefe a colaborar con el IB en la discusión y la mejora de las evaluaciones. La mayoría de los examinadores jefe son también examinadores principales de uno de los componentes de los que son responsables, así que disponen de experiencia directa sobre las respuestas de los alumnos en su área disciplinaria. A diferencia de los examinadores principales, los examinadores jefe también deben velar por la calidad de las pruebas de examen de sus asignaturas durante la fase de creación.

Los examinadores jefe no forman parte del personal del IB, sino que tienen un contrato para trabajar con la organización durante un período de entre dos y siete años. Este límite de siete años existe para alentar la progresión y garantizar que las evaluaciones del IB no se conviertan en algo monótono.

En algunas asignaturas con un número reducido de alumnos, no es adecuado que el IB nombre a ningún examinador jefe. En esos casos, el IB nombrará a un examinador responsable, que asumirá la función de garantizar la coherencia entre componentes y hará recomendaciones al IB sobre los límites de calificación. Esta persona también desempeñará la función de examinador principal para un componente.

Normalmente, los examinadores principales y los examinadores jefe participarían en el ciclo de revisión del currículo, pero esto no se aborda en este documento.

#### Otros examinadores

El examinador es responsable de corregir el trabajo de los alumnos según el estándar que fija el examinador principal. Para garantizar que comprende y aplica dicho estándar, se utiliza el modelo de calidad (véase la sección "Modelo de calidad").

Los examinadores interesados solicitan trabajar con el IB para realizar la corrección en una convocatoria de exámenes. Dichos examinadores deben ser expertos en la asignatura que quieran corregir (normalmente serán profesores de esa asignatura) y deben tener experiencia en la enseñanza de alumnos de la franja de edad correspondiente. El IB se pone en contacto con las personas de referencia que aporta el aspirante a

examinador para comprobar sus credenciales. Por lo general, a los examinadores solamente se les ofrece corregir un componente de evaluación interna (trabajo de clase) y un componente de examen. El período de corrección de esos dos componentes no se solapa. A los examinadores se les paga por cada cuestionario de examen "en condiciones reales". A efectos de la remuneración, los exámenes de aptitud y los de control (véase la sección "Modelo de calidad") no se tienen en cuenta, porque su función es demostrar al IB que los examinadores comprenden el estándar de corrección.

Los jefes de equipo son examinadores con mucha experiencia a los que el IB pide que ayuden a otros examinadores a comprender el estándar de corrección. Los jefes de equipo prestan apoyo a los examinadores durante el proceso de corrección de los exámenes de aptitud y ofrecen comentarios si las pruebas de control indican que se están desviando de los estándares necesarios. El número de jefes de equipo en una asignatura dependerá del número de examinadores que hagan falta. En una asignatura con pocos alumnos, el examinador principal puede prestar apoyo a todos los examinadores.

Los examinadores supervisores son examinadores con experiencia a quienes se pide que ayuden a los examinadores principales en distintas tareas. Normalmente, los examinadores supervisores también son jefes de equipo.

#### Las responsabilidades del personal del IB

Cada asignatura se asignará a un responsable de evaluación, quien gestionará todo el ciclo de evaluación y comprobará que se sigan los procedimientos de control de calidad pertinentes. Muchos responsables de la asignatura también son expertos en la asignatura correspondiente, por lo que pueden ofrecer apoyo a los examinadores jefe y a los examinadores principales.

En el IB, varios equipos gestionan las solicitudes de modificación de cuestionarios de examen, de disposiciones especiales o de consideraciones especiales, las alegaciones de conducta improcedente (incluido el plagio), la supervisión de la calidad del trabajo de los examinadores y el cálculo de los factores de moderación.

El director en jefe de evaluación, con el apoyo del jefe de evaluación del programa y del jefe de principios y prácticas de evaluación, es el responsable de estudiar los límites de calificación que recomienda el examinador jefe de cada asignatura, y decidir si está conforme o pedirle al examinador 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 expertos externos para elaborar evaluaciones justas y de calidad, la responsabilidad final recae sobre el IB.

#### Funciones relacionadas con la creación de las pruebas de examen

Las habilidades clave que debe tener un examinador principal son ser coherente con el estándar de corrección y ser capaz de explicar a otros examinadores en qué consiste el estándar requerido.

Las habilidades clave necesarias para redactar un examen de gran calidad son muy distintas e implican demostrar:

- Creatividad para plantear preguntas interesantes y distintivas que reflejen el currículo.
- Conciencia y comprensión de los distintos enfoques pedagógicos y culturales para elaborar preguntas sin sesgos.
- Habilidades lingüísticas para plantear preguntas claras y sin ambigüedades que conserven estas características cuando se traduzcan a las demás lenguas requeridas.
- Claridad para formular preguntas que permitan evaluar específicamente aquello para lo que están pensadas (pertinencia del constructo), de modo que se puedan corregir de manera coherente (fiabilidad).

En general, los examinadores principales y los examinadores supervisores cuentan con ambos conjuntos de habilidades. Esto permite disfrutar del beneficio de que las mismas personas redacten la evaluación y la corrijan, con la ayuda de los responsables de las asignaturas.



Además de estas funciones creativas, hay otras personas que intervienen en la producción de las evaluaciones finales:

- Diseñadores técnicos: son los responsables de dar a las preguntas de evaluación su formato final, ya sea en papel o en pantalla.
- Traductores: son los encargados de convertir los cuestionarios de examen a las otras lenguas en las que se realizará la evaluación.
- Revisores externos: son expertos en las asignaturas que responden a las preguntas de la evaluación como si fueran alumnos; de esta manera, pueden proporcionar comentarios sobre la extensión total de los cuestionarios de examen y señalar cualquier error o ambigüedad que pueda haber en el texto.

Para obtener más información sobre cómo se elaboran las evaluaciones, consulte la sección "Preparación de pruebas de examen: desarrollo y calidad".

### Jerarquía de los examinadores

- El examinador principal es quien tiene la decisión final sobre la puntuación que se concede a un componente. Los demás examinadores tienen que seguir ese estándar.
- El IB parte de la base de que las personas que han colaborado con el examinador principal para establecer el estándar comprenden mejor dicho estándar que quienes lo han aprendido mediante el material de estandarización y los jefes de equipo. Esto significa que las puntuaciones de los examinadores que están más cercanos al examinador principal en la jerarquía prevalecen sobre aquellas de quienes estén más alejados.
- El adjetivo "supervisor" implica proximidad al examinador principal en la jerarquía de examinadores, y no necesariamente una larga experiencia como examinadores o como docentes.

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

El examinador principal cuenta con el apoyo de un equipo de examinadores supervisores para fijar el estándar. Como estos examinadores tienen la oportunidad de discutir el estándar con el examinador principal, el IB considera que ellos serán quienes mejor lo entiendan. Por ese motivo, si no están de acuerdo con la puntuación que otorque otro examinador (normalmente como resultado de una revisión de la corrección de casos en riesgo, o de una consulta sobre los resultados), el IB utilizará como puntuación final para ese trabajo aquella que otorque el examinador que ocupe un puesto más alto en la jerarquía.

La mayoría de los jefes de equipo son examinadores supervisores, pero cuando no sea así, la puntuación del examinador supervisor estará por delante de la de un jefe de equipo. Del mismo modo, se da preferencia a la puntuación de un jefe de equipo antes que a la de un examinador que ha aprendido el estándar del examinador principal gracias al material de estandarización y a las instrucciones de su jefe de equipo.

Todo esto explica la jerarquía de los examinadores. Para una asignatura que tenga una gran cantidad de alumnos, la estructura de la jerarquía podría ser similar a la que se presenta a continuación.

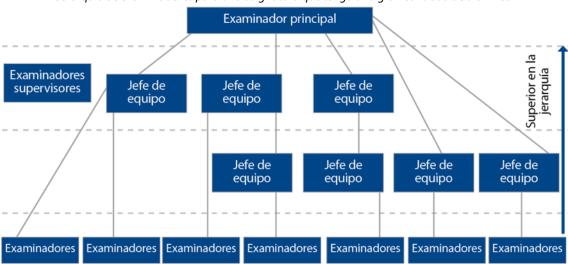


Figura 39 Jerarquía de examinadores para una asignatura que tenga una gran cantidad de alumnos

Esta jerarquía no es rígida. Si los datos procedentes de la corrección de exámenes de control o de aptitud muestran que un examinador comprende el estándar del examinador principal mejor que un miembro con más nivel en la jerarquía, esto se tendrá en cuenta. La función del jefe de equipo no solo exige aplicar el estándar, sino también ser capaz de explicarlo.

En una situación ideal, no haría falta tomar decisiones basándose en la jerarquía, porque todos corregirían los exámenes 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 con las respuestas del alumno se remite al examinador principal para que este determine la puntuación adecuada.



## La integridad de la evaluación

- Las evaluaciones del IB solamente pueden ser justas si se da igualdad de oportunidades a todos los
- Las distintas formas de mala administración y de conducta improcedente ponen en situación de desventaja a los alumnos que han seguido las normas; por eso, el IB hace todo lo posible para evitar tales comportamientos.

Las normas del IB, que se detallan en el Reglamento general y otros documentos de cada programa, están diseñadas para minimizar las posibilidades de que haya conducta improcedente o mala administración. En última instancia, el colegio es el único que puede crear una cultura de aprendizaje en la que la conducta improcedente se rechace y se denuncie.

- Si bien ciertas formas de evaluación son menos vulnerables a la conducta improcedente 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.
- Es muy probable que la evaluación en pantalla sea de gran ayuda para mantener la integridad de las evaluaciones del IB.
- Si alguien sospecha de algún caso de conducta improcedente que el colegio no esté abordando, deberá comunicarse con complaints@ibo.org o con "El IB responde".

La probidad académica es un conjunto de valores y habilidades que promueven la integridad personal y las buenas prácticas en la enseñanza, el aprendizaje y la evaluación. No se debe considerar como una serie de normas estrictas que se tienen que imponer, sino que debe adoptar la forma de una cultura en el seno del colegio y en la comunidad más amplia, incluidos los tutores legales. Si bien resulta fácil explicar a los alumnos qué es la falta de probidad académica haciendo referencia a la colusión, el plagio o conductas ilícitas, este enfoque no logrará originar el tipo de cultura positiva de integridad que, de manera natural, provocará que los resultados de la evaluación sean más justos.

Para que una evaluación sea válida, debe reflejar de manera precisa los logros de un alumno en comparación con el resto de los alumnos que han 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 de la creación de igualdad de condiciones.

En el Reglamento general se define la conducta improcedente como "toda acción [...] de un alumno matriculado por la cual este u otro alumno matriculado sale o puede salir beneficiado injustamente". Tal actividad no afecta solamente a los alumnos que se ven implicados, sino a todos los que han realizado la evaluación al mismo tiempo que ellos, ya que reduce la validez de la evaluación. Por ese motivo, el IB se toma muy en serio la conducta improcedente. En los Procedimientos de evaluación de cada programa y en las publicaciones sobre probidad académica se encuentra información más detallada sobre cómo prevenir la conducta improcedente y sobre sus consecuencias.

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

Sin embargo, la manejabilidad es otro aspecto importante de la validez, y se debe tener en consideración al tratar de impedir que haya conductas improcedentes. La preparación de las salas de examen, en particular en el caso de la evaluación en pantalla, 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 los coordinadores y los directores de colegio le comuniquen qué prácticas han encontrado óptimas y cuáles les han resultado complicadas.

Esta comunicación también permite comprender mejor nuevas cuestiones que surgen en lo relativo a la probidad académica, y todo aquel que tenga alguna inquietud o reflexión al respecto puede escribir a complaints@ibo.org o a "El IB responde".

## La gestión de los conflictos de interés

En el IB se controla minuciosamente el acceso a los cuestionarios de examen, y la organización gestiona de manera activa cualquier conexión con alumnos 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 puede corregir el trabajo de sus propios alumnos. 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 independiente revisaría la corrección para determinar que nada indica que se haya aplicado un estándar distinto a los propios alumnos del examinador.

# La gestión de la mala administración

En cada programa del IB, el *Reglamento general* y los *Procedimientos de evaluación* establecen las normas e instrucciones para minimizar el riesgo de mala administración.

Los cuestionarios de examen se deben conservar en una caja de seguridad cerrada con llave para evitar que accedan a ellos personas no autorizadas. Los exámenes se envían en bolsas a prueba de manipulación que permiten ver inmediatamente si se han abierto. Las filtraciones de cuestionarios de examen conforman una de las situaciones más complicadas de evitar para el IB, ya que es muy complicado saber en qué medida se han hecho circular y se han compartido. Por ese motivo, la acción que se lleve a cabo normalmente afectará a un gran número de alumnos, para garantizar que la experiencia sea la misma para todos.

Los paquetes con los cuestionarios de examen se deben abrir delante de los alumnos, ya que de esta manera verán que los exámenes se han guardado de manera segura. En caso de que no suceda así, los alumnos deben ponerse en contacto con el IB.

Los supervisores de exámenes deben vigilar durante todo el examen para asegurarse de que no se produzca ninguna conducta improcedente.

El IB realiza inspecciones en los colegios durante las convocatorias de exámenes para comprobar que esas prácticas se llevan a cabo. A pesar de todo, estos son controles esporádicos de los procesos, y el IB otorga una gran responsabilidad a los directores de colegio y a los coordinadores de los programas para evitar situaciones de mala administración, ya que estos profesionales son quienes están en la mejor posición para garantizar que las normas de calidad se mantengan en sus colegios día a día. Además, ellos deben encargarse de que la cultura escolar de su colegio aliente a sus docentes y a sus alumnos a adoptar las mejores prácticas y la mayor integridad posibles.

# Dificultades provocadas por los exámenes internacionales y las zonas horarias

El IB tiene colegios en todo el mundo, lo cual implica algunas dificultades específicas. En la mayoría de los sistemas nacionales, todos los alumnos realizan los exámenes al mismo tiempo. Sin embargo, el IB no puede adoptar esta práctica sin poner en situación de desventaja a algunos alumnos, debido a las distintas zonas horarias. Imaginemos, por ejemplo, si un examen comenzara a las cinco de la mañana o que terminara pasada la medianoche.



Esta realidad implica que algunos alumnos habrán acabado sus exámenes antes de que otros los hayan empezado. Por ese motivo, se requiere a los alumnos y a los profesores que sean íntegros para no beneficiarse de esta situación. En los casos en que hay suficientes alumnos, se crean dos cuestionarios de examen distintos en función de las zonas horarias, con el fin de reducir el tiempo entre el fin de un examen para los alumnos de una zona y el inicio del mismo examen para los alumnos de la otra zona. Sin embargo, incluso en estos casos puede haber una considerable variación en la hora de comienzo de un examen.

Existen varias soluciones para atenuar esta situación. La primera de ellas es el interés propio: ¿por qué motivo los alumnos elegirían beneficiar a otros alumnos? La segunda es supervisar minuciosamente los sitios web en los que podrían compartirse los cuestionarios de examen. La regla que prohíbe a los alumnos sacar los cuestionarios de examen de la sala pretende precisamente limitar las posibilidades de compartir en línea las preguntas.

La tercera es asegurarse de que las preguntas estén diseñadas para evaluar la comprensión más que la capacidad de recordar conocimientos. Dada la limitación de los plazos de tiempo que hay, no resulta una gran ventaja contar con un poco más de tiempo para pensar en la respuesta si el examen no evalúa la rapidez de pensamiento (cosa que sí sucede en los exámenes que miden la capacidad de recordar conocimientos). La última solución que se utiliza es prohibir a los alumnos 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.

# El plagio de los trabajos de clase

Una de las áreas más difíciles de gestionar es asegurarse de que el trabajo que el alumno realiza en clase (y en casa) sea obra suya.

El IB es consciente de la existencia de numerosos sitios web que ofrecen "ayuda" a los alumnos con su trabajo. La mejor defensa contra este tipo de conducta improcedente es contar con el profesor, que es quien ha trabajado con el alumno y puede identificar si un trabajo no se corresponde con el nivel habitual del alumno. Por este motivo, se pide tanto al alumno como al profesor que confirmen que el trabajo que se entrega al IB es obra del alumno. Esto no es un simple requisito más: el IB no está obligado a aceptar otras versiones una vez que se ha entregado el trabajo.

Todos los Colegios del Mundo del IB deben cumplir el requisito de establecer y gestionar una cultura de probidad 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. No basta con utilizar el software Turnitin: los profesores también deben trabajar con los alumnos cuando estos redacten sus trabajos de evaluación interna, con el fin de comprobar la autenticidad de dichos trabajos.

A fin de evitar que haya dudas en este respecto, no se acepta ningún nivel de plagio, y todas las citas utilizadas deben documentarse adecuadamente tal como establecen las normas del IB. Para detectar casos de plagio, el IB utiliza una variedad de *software*, entre ellos Turnitin.

## Ventajas de la evaluación en pantalla

La evaluación en pantalla ofrece varias ventajas para gestionar la conducta improcedente. Este sistema limita las posibilidades de acceder a los cuestionarios antes del examen, ya que se utilizan datos cifrados y contraseñas. Además, permite controlar cuándo y quién abre los paquetes de examen.

Desde la perspectiva de la mala administración, el IB puede exigir a los colegios registrar y justificar cualquier modificación que realicen, como añadir tiempo o pausas a un examen. 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.

El sistema electrónico también permite saber con más detalle en qué momento se dieron las respuestas. Esto posibilita, por ejemplo, demostrar que un alumno respondió a una pregunta antes de que ocurriera un incidente. Además, permite comprobar a gran escala la existencia de respuestas similares, algo que no es posible con los exámenes manuscritos.

El IB es consciente de que las evaluaciones en pantalla se enfrentarán a nuevas formas de conducta improcedente, en particular a intentos de piratería informática, pero ya se están desarrollando herramientas para repeler dichos ataques.

#### Recursos

Los profesores, los colegios y los alumnos son quienes se encuentran en la mejor posición para alzarse contra los casos de conducta improcedente y detenerlos, ya que pueden crear una cultura en la que dichas conductas sean inaceptables, y también prestar atención para atajar estas situaciones cuando ocurran. Los materiales que se presentan a continuación pueden ayudar a los colegios a crear una cultura de integridad académica. Dichos materiales están disponibles en el sitio web del IB, en la sección "Herramientas digitales".

- La probidad académica en el contexto educativo del IB
- Uso eficaz de citas y referencias

Si desea recibir más apoyo, o si sospecha de algún caso de conducta improcedente que el colegio no esté abordando, comuníquese con complaints@ibo.org o con "El IB responde".



# Equidad para todos: cómo satisfacer las necesidades de los alumnos

Para que las evaluaciones del IB sean válidas, no pueden discriminar a los alumnos con necesidades especiales. Por ello, el IB tendrá en cuenta las peticiones de ofrecer exámenes modificados y adecuaciones inclusivas, tal como se establece en los Procedimientos de evaluación de cada programa.

- La mejor manera de garantizar la equidad es diseñar evaluaciones que sean accesibles para todos, lo cual eliminaría la necesidad de realizar modificaciones. Esta idea queda englobada en el concepto de diseño universal de evaluación que forma parte del compromiso del IB con el diseño universal para el aprendizaje (DUA).
- En ocasiones, las necesidades de algunos alumnos se conocen previamente y se les da respuesta con adecuaciones inclusivas de acceso, entre ellas el uso de cuestionarios de examen modificados.
- Sin embargo, a veces hay circunstancias que surgen en el último momento o que no se pueden solucionar mediante adecuaciones inclusivas. En dichos casos se aplican los procesos de consideración especial.
- En última instancia, el propósito de estas adecuaciones es que haya equidad para todos los alumnos. En consecuencia, a la hora de tomar decisiones, el IB debe tener en cuenta lo que es justo para todo el grupo y no solo para un alumno concreto. El objetivo es que todos los alumnos estén en igualdad de condiciones.

El IB cree que todos los alumnos deben tener la posibilidad de demostrar sus capacidades en condiciones de examen lo más equitativas posible. Las condiciones normales de evaluación pueden representar una desventaja para aquellos alumnos con necesidades de apoyo para el aprendizaje, impidiéndoles demostrar su nivel de conocimiento y capacidad. Asimismo, el IB es consciente de que, en ocasiones, hay circunstancias o hechos que escapan al control de los alumnos y que pueden afectar a su desempeño, por lo que se deberían tener en cuenta.

La mejor forma de garantizar la equidad en una evaluación es que todo el mundo realice la misma evaluación de la misma manera. Muchas de las modificaciones que se hacen para cubrir las necesidades específicas servirían para ayudar a todos los alumnos a entender y responder las preguntas. La situación ideal es que todas las evaluaciones se desarrollen teniendo en cuenta la variedad de necesidades que pueden tener los alumnos, en vez de tratar a algunos alumnos de manera distinta. En eso consiste el concepto de diseño universal de evaluación. El IB es consciente de que a veces no es posible adoptar este enfoque de inclusividad total, y por eso también cuenta con un proceso para solicitar adecuaciones inclusivas específicas.

Las adecuaciones inclusivas de acceso están diseñadas para satisfacer las necesidades específicas de los alumnos, por ejemplo, aquellos que tienen:

- Dificultades de aprendizaje
- Dificultades del lenguaje
- Dificultades de aprendizaje específicas
- Dificultades de comunicación o del habla
- Trastornos del espectro autista
- Dificultades sociales, emocionales o de conducta
- Discapacidades varias o trastornos motores o sensoriales, afecciones médicas o problemas de salud mental

Se considerará la realización de ajustes razonables para ciertos alumnos con el fin de satisfacer sus necesidades específicas. Para obtener más información, consulte los Procedimientos de evaluación del programa correspondiente y la publicación del IB Alumnos con necesidades específicas de acceso a la evaluación.

**Circunstancias adversas** son todas aquellas circunstancias ajenas a la voluntad del alumno que puedan perjudicar su desempeño. El Comité de la evaluación final valorará dichas circunstancias y, en caso de decisión favorable, los alumnos que estén cerca de un límite de calificación recibirán la calificación superior. El principio de equidad para todos los alumnos, que defiende el IB, implica que, al considerar cualquier adecuación inclusiva o circunstancia adversa, no se debe crear una situación injusta para los otros alumnos

# Principios para las adecuaciones inclusivas de acceso

que realicen las evaluaciones. El objetivo es que todos los alumnos estén en igualdad de condiciones.

Los principios para las adecuaciones inclusivas de acceso se establecen en la publicación del IB *Alumnos con necesidades específicas de acceso a la evaluación*. El texto que aparece a continuación se ha extraído de la versión del PD, pero en todos los programas se aplican principios similares.

- El IB debe garantizar que la calificación final concedida a un alumno en cualquier asignatura no sea una descripción equívoca de su nivel de conocimiento y capacidad. Por tanto, se aplican los mismos estándares de evaluación para todos los alumnos, independientemente de si tienen o no necesidades de apoyo para el aprendizaje.
- 2. Las adecuaciones inclusivas de acceso, que incluyen ajustes razonables, son medidas que se adoptan antes de un examen para que un alumno pueda acceder a la evaluación. Dichas adecuaciones no pueden solicitarse con carácter retrospectivo, ya sea para un examen oral o escrito.
- 3. Las adecuaciones solicitadas para un alumno no deben representar una ventaja para dicho alumno en ningún componente de evaluación.
- 4. Las adecuaciones inclusivas de acceso descritas en este documento han sido concebidas para alumnos con un nivel de aptitud suficiente como para satisfacer los requisitos de evaluación necesarios para que se les otorque el diploma o los resultados de los cursos.
- 5. Cuando sea necesario, el colegio puede proporcionar adecuaciones inclusivas de acceso para un alumno, ya sea durante el curso de sus estudios en el Programa del Diploma o en exámenes de práctica. En este documento se enumeran las adecuaciones que pueden adoptarse para la evaluación sin que sea necesario solicitar la autorización del IB; en los demás casos, será obligatorio obtener autorización del centro global del IB en Cardiff. De igual modo, si un alumno del Programa del Diploma tiene problemas para satisfacer los requisitos de Creatividad, Actividad y Servicio (CAS), se deberá consultar escribiendo a "El IB responde".
- 6. Se recomienda a los colegios que planifiquen las adecuaciones inclusivas de acceso para sus alumnos basándose en los criterios del IB, tal como se establece en esta política y en las observaciones de los profesores del alumno en el aula durante las pruebas y el trabajo en clase.
- 7. Las adecuaciones inclusivas de acceso que sean solicitadas para un alumno deben ser las que tenga normalmente para trabajar durante el programa de estudios. El IB autorizará únicamente en casos muy poco comunes y realmente excepcionales aquellas solicitudes de adecuaciones inclusivas de acceso que no sean el modo normal de trabajo del alumno y que solo se hayan realizado para respaldar al alumno en algún momento de los seis meses anteriores a la petición, justo antes de los exámenes.
- 8. El IB tiene como propósito autorizar las adecuaciones inclusivas de acceso que sean compatibles con aquellas que normalmente están a disposición del alumno en cuestión. No obstante, solo se autorizarán las adecuaciones que sean coherentes con la normativa y las prácticas del IB. No se debe dar por supuesto que el IB admitirá las adecuaciones solicitadas por los colegios. Los coordinadores deben proporcionar información sobre el método habitual de trabajo del alumno en el aula.
- La filosofía educativa del IB se basa en la mentalidad internacional. Esto implica que su normativa de adecuaciones inclusivas de acceso puede no reflejar las prácticas habituales de un país en particular.



- Esta normativa es resultado de un estudio de las prácticas establecidas en diferentes países, con el objeto de lograr la equidad entre alumnos con necesidades específicas de acceso a la evaluación.
- 10. El IB se asegurará, siempre que sea posible, de que se aplican las mismas disposiciones para alumnos con similares necesidades específicas de acceso. Debido a las diferencias culturales que existen con respecto a la identificación de las necesidades de apoyo para el aprendizaje y a la naturaleza de las adecuaciones de acceso autorizadas en los colegios, un cierto grado de flexibilidad puede ser necesario para garantizar la equiparación entre alumnos de diferentes países.
- 11. 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 haya solicitado el coordinador.
- 12. El IB trata toda la información sobre los alumnos de forma confidencial. Solo podrán acceder a dicha información, de ser necesario, el personal habilitado del IB y los miembros del Comité de la evaluación final, a quienes se les pide que respeten su carácter confidencial.
- 13. Si un colegio no satisface las condiciones especificadas por el IB a la hora de administrar las adecuaciones inclusivas de acceso, o introduce disposiciones especiales sin autorización, es posible que al alumno no se le conceda una calificación final en la asignatura y nivel en cuestión.
- 14. Si se puede demostrar que la falta de dominio de la lengua de respuesta de un alumno es consecuencia de una necesidad de apoyo para el aprendizaje que ha sido identificada, se pueden autorizar adecuaciones inclusivas de acceso. (Para las asignaturas en los grupos 3 a 6, todos los alumnos pueden usar un diccionario bilingüe en los exámenes escritos).
- 15. Los colegios no deben informar a los examinadores de las circunstancias adversas o la situación particular de un alumno (como autismo, dificultades para escribir, etc.).
- 16. En el caso de los trabajos evaluados internamente, los profesores no deben hacer ningún ajuste al calificar el trabajo del alumno.
- 17. La lista de las adecuaciones inclusivas de acceso permitidas se revisa regularmente. El IB tomará en consideración adecuaciones alternativas propuestas por los coordinadores, siempre y cuando todos los alumnos con necesidades similares puedan beneficiarse de dichas adecuaciones.
- 18. De acuerdo con el documento Reglamento general del Programa del Diploma, los alumnos que cursan este programa pueden participar en tres convocatorias de exámenes para obtener el diploma. A criterio del IB, un alumno con necesidades de apoyo para el aprendizaje podrá participar en convocatorias adicionales.
- 19. Si la naturaleza de la dificultad de un alumno o de la adecuación inclusiva de acceso que se haya autorizado pudieran causar molestias a otros alumnos durante un examen, el alumno deberá realizar el examen en una sala independiente y ser supervisado de acuerdo con el reglamento que rige la realización de los exámenes del Programa del Diploma.
- 20. Los exámenes escritos deben ser supervisados de acuerdo con el reglamento que rige la realización de los exámenes del Programa del Diploma. El supervisor de examen no deberá ser un familiar del alumno ni ninguna persona con la que pueda existir un conflicto de intereses evidente o aparente.
- 21. 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 durante los exámenes deberán comunicarse a "El IB responde" lo antes posible.

#### Exenciones de evaluación

Normalmente no se autorizan exenciones respecto de los componentes de evaluación. No obstante, si un componente o una parte de la evaluación exige una función fisiológica que un alumno 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 todas las alternativas posibles. Solo se autorizará una exención cuando existan motivos importantes. La incapacidad física del alumno 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 el documento *Alumnos con necesidades específicas de acceso a la evaluación* del programa correspondiente.

# Oportunidades para adecuaciones inclusivas con la evaluación en pantalla

El uso de la evaluación en pantalla permite que el alumno tenga más control sobre el modo en que le gustaría que se presentara la evaluación. Gracias a las computadoras, las evaluaciones se pueden presentar en distintos tipos de fuente, tamaños de texto y colores para satisfacer las necesidades de los alumnos. Este tipo de ajustes se puede poner fácilmente a disposición de todos los alumnos.

Actualmente, en muchos casos la adecuación de inclusión que se solicita es permitir utilizar una computadora. Esta necesidad se ve claramente satisfecha con la evaluación electrónica, siempre y cuando el *software* requerido que permita la inclusión sea compatible con la herramienta en pantalla. El IB es muy consciente de este requisito y procura que todas las evaluaciones en pantalla cumplan con los estándares habituales en el sector en lo referido al *software* que posibilita la inclusión.

#### Circunstancias adversas

Circunstancias adversas o imprevistas son todas aquellas circunstancias ajenas a la voluntad del alumno 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. 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 todos los alumnos cumplan 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, consulte los *Procedimientos de evaluación* y el *Reglamento general* del programa correspondiente.

En esos casos, el Comité de la evaluación final estudiará las pruebas que aporte el colegio, con objeto de determinar si el alumno afectado puede tener derecho a consideraciones especiales. Si las circunstancias se consideran "adversas" y dignas de "consideración", se podrá realizar un ajuste de la puntuación total del alumno para las asignaturas o los componentes troncales del programa en cuestión. Si el alumno se encuentra a uno o dos puntos del límite de calificación superior para la asignatura o las asignaturas en cuestión, la calificación se aumentará.

#### Diseño universal de evaluación

En los párrafos anteriores se ha explicado cómo gestiona el IB las situaciones en las que los alumnos necesitan modificaciones o ayuda específica para poder realizar de manera justa las evaluaciones. A pesar de todo, la mejor solución es que las evaluaciones no supongan ninguna barrera para los alumnos. El concepto de diseño universal de evaluación se refiere a plantearse el acceso, la inclusión, la igualdad, las sensibilidades culturales, los estereotipos y los sesgos desde el primer diseño de la evaluación. Esto incluye la creación de las tareas y las preguntas de examen, y además aborda el diseño del modelo general de evaluación que determina cómo crear evaluaciones equiparables para cada convocatoria de exámenes. Si se diseñan evaluaciones más inclusivas desde el principio, no solo aumentará la pertinencia del constructo, sino que también se minimiza el riesgo de no satisfacer las necesidades de todos los alumnos.

El diseño universal de evaluación es un aspecto del concepto más general de diseño universal para el aprendizaje (DUA). El DUA procura crear entornos de aprendizaje accesibles para todos los alumnos, incluidos aquellos con discapacidades, de diversas procedencias culturales y lingüísticas, superdotados y talentosos. Los principios del DUA se pueden aplicar en todas las áreas de la educación, en particular en el



diseño del currículo, en la gestión escolar y en la enseñanza. Para obtener más información sobre el DUA en el contexto del IB, consulte: Rao, K.; Currie-Rubin, R. y Logli, C. UDL and Inclusive Practices in IB Schools Worldwide, 2016.

### El ciclo de evaluación

- El proceso de creación de las evaluaciones debe imaginarse como un ciclo continuo en el que cada paso viene formado por el anterior y sirve de base para el siguiente.
- De media, se tarda 18 meses en preparar un cuestionario de 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 los alumnos, corrección por parte de los examinadores y publicación de los resultados. Lo importante de todo este proceso es que de cada convocatoria de exámenes se obtienen lecciones para mejorar la calidad de la evaluación en la siguiente convocatoria.

El siguiente diagrama 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 que hay después del diagrama se resume cada etapa y se ofrece un enlace a la sección pertinente.

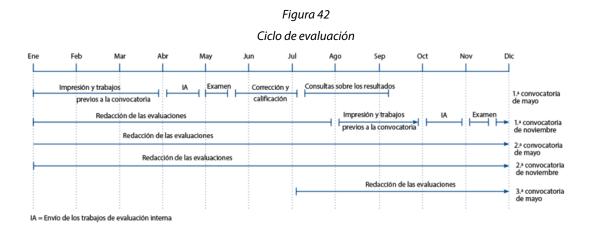


Figura 41 El ciclo de evaluación

Etapa del ciclo de evaluación	Descripción
Preparación del examen	Proceso de creación de cada examen. Cubre un amplio espectro: decidir los temas que tratarán las preguntas, redactar y corregir las preguntas

Etapa del ciclo de evaluación	Descripción
	concretas, traducir el examen a otras lenguas, darle el formato deseado y, finalmente, realizar los controles de calidad necesarios.
Exámenes	Parte del proceso en la que los alumnos realizan las evaluaciones en los colegios.
Fijación del estándar	Proceso en el que los examinadores supervisores explican a sus equipos de examinadores cómo corregir los trabajos de los alumnos, e identifican los exámenes "corregidos definitivamente" que se utilizarán para comprobar la calidad del trabajo de los examinadores.
Corrección	Parte en la que los examinadores revisan el trabajo de cada alumno y deciden qué puntuación le otorgan. Los examinadores deben seguir las instrucciones del examinador principal, y la calidad de sus correcciones se comprueba mediante controles regulares.
Concesión de calificaciones	Etapa en la que los examinadores supervisores deciden cómo convertir las puntuaciones (que dependen del examen) en calificaciones finales (que siempre significan lo mismo, independientemente del examen).
Revisión de la corrección de casos en riesgo	El control de calidad final de la corrección. Se centra en aquellas áreas en las que hay pruebas de que puede haber algún problema.
Publicación de los resultados	Difusión de los resultados para los colegios y los alumnos. En esta etapa se incluyen también procesos principalmente administrativos como determinar si los resultados del curso significan que el alumno ha superado el programa, y enviar los certificados de calificaciones finales a las universidades.
Consulta 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.
Diseño de la evaluación	Este es el aspecto más importante para cerrar el ciclo de evaluación. Se trata de aprender de las experiencias de los alumnos que realizan los exámenes y de mejorar lo que se quiere evaluar y el enfoque adoptado, incluidos la cantidad y el tipo de tareas de evaluación.

El siguiente diagrama sirve para tener una idea de cuánto se tarda en desarrollar un examen y del tiempo dedicado a cada parte del ciclo de evaluación. El ciclo representa 3 años completos y destaca que, de media, se tarda 18 meses en preparar un cuestionario de examen. También muestra que el IB prepara al mismo tiempo los cuestionarios de examen de la convocatoria de mayo y noviembre y los del año siguiente.



# El impacto de la evaluación electrónica en el ciclo de evaluación

La introducción de la evaluación electrónica 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 y los exámenes escritos de los alumnos a centros de escaneado). A pesar de todo, habrá que seguir completando cada una de las partes del ciclo de evaluación.



# Preparación de pruebas de examen: desarrollo y calidad

- Esta etapa conduce a cuestionarios de examen completos que están listos para los alumnos.
- Los cuestionarios de examen deben pasar por distintas etapas de aprobación durante la preparación:

Contenido (redacción de las pruebas de examen)

**Formato** 

Usabilidad

Traducción

Envío a los colegios

• En esta etapa también se tienen en cuenta las modificaciones que soliciten los alumnos (adecuaciones de acceso).

# Reglas que rigen la redacción de los cuestionarios de examen

En la sección sobre los principios de evaluación se describen las cualidades de una buena evaluación, en las que se basa el IB a la hora de redactar los cuestionarios de examen. Para este trabajo, también es importante no perder de vista el propósito de la evaluación. Las seis reglas generales que se presentan a continuación se pueden considerar los elementos clave que se espera que los autores de los exámenes tengan en cuenta.

- La autenticidad y la pertinencia del constructo son los aspectos más importantes de la validez.
- La evaluación debe proporcionar a los alumnos oportunidades para que demuestren lo que son capaces de hacer, en vez de servir para identificar lo que no saben hacer.
- Las evaluaciones deben ser accesibles para la mayor cantidad posible de alumnos, pero también deben poder permitir diferenciarlos razonablemente.
- Las mejores evaluaciones son las que resultan accesibles para todos, y el objetivo del IB es que no haga falta realizar adecuaciones inclusivas (diseño universal de evaluación).
- 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.

Concretamente, esto significa que el esquema de calificación se debe desarrollar al mismo tiempo que los cuestionarios de examen, y que hay que incluir orientaciones para los examinadores sobre los márgenes de tolerancia previstos.

#### Términos de instrucción

Los términos de instrucción son palabras y expresiones clave utilizadas en los contenidos de los programas 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.

Aunque los términos de instrucción tienen un significado común en todas las asignaturas, una parte importante de la edición de cuestionarios de examen consiste en asegurarse de que estos términos se utilizan tal como se describe en las guías de las asignaturas del IB pertinentes, ya que los alumnos deben desarrollar su comprensión de estos términos clave en el contexto de cada asignatura.

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 todas las lenguas de respuesta y explica su significado. Esto a veces provoca que se utilicen determinadas palabras de manera poco habitual desde el punto de vista lingüístico, pero es menos problemático que encontrarse con ambigüedades semánticas en las evaluaciones. Al revisar los cuestionarios de examen traducidos, una de las tareas clave es comprobar que se utilicen los términos de instrucción correctos.

Los términos de instrucción están inspirados en gran medida por taxonomías establecidas, como la taxonomía de objetivos educativos de Bloom (Bloom *et al.*, 1956).

#### **Bancos de preguntas**

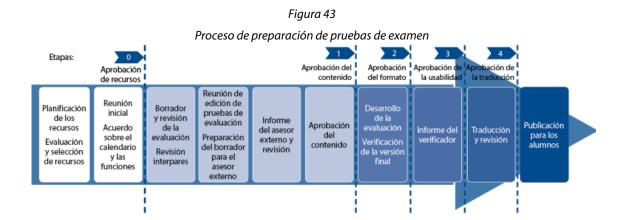
Actualmente, el IB no utiliza ningún tipo de bancos de preguntas. Los autores de los cuestionarios de examen elaboran cada uno de ellos como una evaluación holística.

En otros sistemas de evaluación, las preguntas se ponen a prueba exhaustivamente con el objeto de determinar su exigencia cognitiva, su eficacia como elemento diferenciador y otros aspectos. Toda esta información, así como otros datos más descriptivos (por ejemplo, la relación entre la pregunta y el currículo, la cantidad de puntos, etc.), queda almacenada en el banco de preguntas.

El IB no pone a prueba sus evaluaciones de esta manera, por temor a que se filtren las preguntas y se hagan de dominio público. Además, dada su naturaleza, la mayoría de las preguntas del IB no se pueden poner a prueba sin contar con un grupo de alumnos muy amplio.

Es posible que en el futuro el IB cree su propio banco de preguntas, y tal vez recurra a la Red de educadores del IB (IBEN) para que hagan sus aportaciones. Esto se haría respetando el principio de hacer hincapié en evaluaciones auténticas y pertinentes al constructo.

# Descripción general de la preparación de pruebas de examen



Este diagrama ilustra las distintas etapas del proceso de preparación 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 especialistas externos y personal del IB, como autores de cuestionarios de examen, especialistas en evaluación, personal dedicado a la edición de los exámenes, verificadores externos y miembros del servicio de traducción.

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 de los recursos

La primera etapa del proceso es determinar a las personas que se encargarán de redactar las pruebas de examen. Normalmente esto se encarga a los examinadores 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. Lo normal es que haya un autor por cada prueba de examen. Una vez creada, cada prueba se somete a la revisión de los demás autores de pruebas de examen de la asignatura. Para aquellas asignaturas de lengua con muy pocos alumnos, es posible que una sola persona sea la autora de todas las pruebas de examen.

El siguiente paso es la reunión inicial, en la que todos los participantes en el proceso se reúnen para acordar la responsabilidad asociada a cada tarea y los plazos. En esta reunión intervienen:

- El examinador jefe, que también puede ser uno de los autores del examen (miembro externo)
- El autor de cada prueba de examen de la asignatura (normalmente entre tres y cinco en las asignaturas del Programa del Diploma) (miembro externo)
- El asesor externo (miembro externo)
- El verificador (miembro externo)
- El editor de producción (miembro del personal del IB)
- El responsable de la asignatura (miembro del personal del IB)

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

#### El proceso de aprobación del contenido

Después de la reunión inicial, cada autor debe crear una versión preliminar de su prueba de examen y de su esquema de calificación. Los autores de las otras pruebas de examen de la asignatura revisan estos materiales. Más tarde, se celebra una reunión de edición de pruebas de evaluación, en la que cada prueba de examen se revisa para comprobar lo siguiente:

- El nivel de exigencia cognitiva de las preguntas, y el grado de dificultad general la prueba
- La existencia de posibles sesgos
- Si la prueba de examen sigue las mejores prácticas de accesibilidad
- En qué medida las preguntas cubren el currículo
- La correspondencia entre la prueba de examen y el modelo de evaluación publicado
- Cómo es la nueva prueba de examen en comparación con pruebas y materiales de muestra anteriores
- El posible uso de preguntas que ya se hayan empleado como ejemplo en algún recurso publicado

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

El asesor externo es un experto en la asignatura que aún no ha participado en la redacción de las pruebas de examen ni de los esquemas de calificación. Este aporta una mirada fresca a lo que se ha creado hasta el momento, y debe comentar el nivel de dificultad de las pruebas de examen en conjunto, cómo son en comparación con exámenes anteriores y en qué medida se corresponden con el currículo. El asesor externo propone cambios que deberán revisar el autor y el responsable de la asignatura.

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

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

#### El proceso de aprobación del formato

Una vez aprobado el contenido, el texto de las preguntas no se puede modificar. A continuación, el editor de producción del IB les dará a las preguntas el formato que verán los alumnos en el examen.

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 para imprimirlos.

Para la evaluación en pantalla, el editor de producción tendrá que crear los exámenes en un entorno específico denominado *Sistema de desarrollo de la evaluación electrónica del IB.* En él se incluyen las instrucciones y los estilos adecuados.

En ambos casos, las pruebas de examen completadas se comparan con el contenido aprobado.

#### El proceso de aprobación de la usabilidad

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

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

En algunas asignaturas de lengua con muy pocos alumnos, la función del verificador la realiza un segundo hablante nativo de la lengua en vez de un experto en literatura.

Después de considerar los comentarios del verificador, las pruebas de evaluación se finalizan. Llegados a este punto, ya están listas para imprimirse (en el caso de las evaluaciones en papel) o cargarse en el sistema (en el caso de las evaluaciones en pantalla) 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 los alumnos que realizan los exámenes. El IB se toma este tema muy seriamente.

En los procesos descritos anteriormente se ve que en cada etapa existe un control de la calidad además de la aprobación oficial, de la que se encargan el asesor externo, el revisor o el verificador. La experiencia indica que es mejor no realizar varias comprobaciones en busca del mismo error: los controles de calidad son generalmente menos escrupulosos si se sabe que otra persona realizará una comprobación posterior.

En la producción de pruebas de examen, el IB se adhiere al principio de "hacerlo bien a la primera y revisarlo una sola vez".

## **Traducciones**

Los alumnos deben poder realizar los exámenes en el idioma en el que se les haya enseñado. Por este motivo, el IB da 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, japonés u otras lenguas.

La traducción de las evaluaciones no es tarea fácil. Es fundamental que el proceso de traducción no altere el significado de las preguntas, para evitar que haya alumnos que salgan favorecidos o desfavorecidos al realizar la evaluación en una lengua concreta. Esto plantea dificultades, ya que hay que reformular las oraciones cuando se traducen a otra lengua. Además, determinadas palabras pueden tener significados distintos en otras lenguas, así que en la traducción se podrían crear nuevos significados o eliminar algunos.

La necesidad de traducir los exámenes se tiene en cuenta desde los primeros borradores, y los equipos encargados de la creación de pruebas de examen son muy conscientes de los problemas que pueden surgir. La traducción formal se realiza al final del proceso, después de la aprobación de la usabilidad. Los traductores que contrata el IB son especialistas en la asignatura, para asegurarse de que el lenguaje técnico que se utiliza en la evaluación sea preciso. Una vez entregada, un asesor bilingüe compara la traducción con el original (y no solamente el texto) para hacer un último control de calidad.



# Cómo gestionar las solicitudes de pruebas de examen modificadas

Lo ideal es diseñar las evaluaciones de modo que resulten adecuadas para todos los alumnos. Sin embargo, hay necesidades y requisitos que no se pueden satisfacer con una mejora del diseño de la prueba de examen. Algunas modificaciones (como utilizar una fuente distinta, un fondo de otro color o un formato mayor) son relativamente fáciles de implementar y se llevan a cabo justo antes de imprimirse las pruebas de examen.

Otras modificaciones, como las pruebas de examen en formato braille, requieren una mayor preparación. También es posible que requieran formular tareas relacionadas pero distintas, para obtener pruebas del grado de dominio del alumno en un área. Un ejemplo de esto sería una pregunta en la que se pidiera comentar una imagen, lo cual no sería adecuado para un alumno ciego. En este caso, habría que plantearse si ofrecer una descripción de la imagen sería una alternativa razonable o si, por el contrario, esto aumentaría la carga cognitiva del alumno (demasiada información) o le proporcionaría demasiada orientación (si la descripción se centra en los aspectos que se valorarán para otorgar la puntuación). Una alternativa sería plantearse si la pregunta se podría formular de otro modo.

El autor de la prueba de examen es quien toma estas decisiones, con la orientación del responsable de acceso e inclusión del IB. Si corresponde, también se consulta a expertos externos. Aunque las decisiones se toman para cada caso concreto, los resultados y las razones de estas se registran sistemáticamente para que se puedan realizar controles de calidad con el fin de asegurar una coherencia. Cuando es necesario realizar modificaciones especiales, estas deben hacerse de modo que las diferencias en aquello que se evalúa sean mínimas.

Estas decisiones se debaten en profundidad, así que lo ideal sería tomarlas antes de la etapa de aprobación del contenido. De ese modo, se garantiza que el autor tenga presentes los propósitos originales de las preguntas, y también se da la posibilidad de mejorar la pregunta original y así evitar que haga falta modificarla. Para ello, los colegios deben informar al IB de las solicitudes de acceso inclusivo lo antes posible.

## El paso a la evaluación en pantalla

La creación de evaluaciones en pantalla provoca algunas diferencias considerables en el modo en que se preparan los exámenes, pero el proceso general que se ilustra en el diagrama continúa siendo el mismo. De hecho, en las evaluaciones en pantalla del PAI ya se sigue este proceso.

El cambio más destacable es la manera en que se desarrollan las pruebas de examen. A medida que se introduzca la evaluación en pantalla, los autores tendrán que comprender las posibilidades que ofrece, con el fin de utilizar óptimamente las herramientas que habrá disponibles. Uno de los aspectos más emocionantes de la evaluación en pantalla es el cambio radical en qué se puede evaluar y cómo, pero sin duda también es una de las cuestiones que presenta mayores dificultades.

Los editores de producción que trabajen en las evaluaciones en pantalla necesitarán una serie de habilidades distintas y más amplias. Es posible que no baste con copiar texto en la plantilla de examen y comprobar los requisitos de estilo, sino que tengan que desarrollar materiales multimedia y asegurarse de que la evaluación en conjunto funcione tal como tenían pensado los autores. El Sistema de desarrollo de la evaluación electrónica del IB que se utiliza actualmente para el PAI ofrece un marco para la evaluación electrónica que es similar al modelo tradicional, pero con una serie de requisitos especiales que los editores deben dominar.

La diferencia más importante será el modo en que los colegios reciben las evaluaciones. Para obtener más información, consulte la Guía para la convocatoria de exámenes del PAI.

#### Exámenes

- El propósito de los exámenes es permitir que los alumnos demuestren todo su potencial en un entorno controlado que ofrezca una experiencia coherente a todos los alumnos y que minimice los riesgos de conducta improcedente.
- Durante la preparación de las convocatorias de exámenes, el IB tiene que hallar un equilibrio entre las necesidades de todos los alumnos y la manejabilidad para los colegios, y finalizar la corrección con la máxima rapidez para cumplir con las expectativas de los alumnos.
- El IB publica reglas claras sobre el comportamiento en la sala de 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 es muy estresante y exigente para los alumnos. Con el fin de reducir la presión a la que se ven sometidos los alumnos, 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, y evitar que haya conductas improcedentes.

# Preparación y gestión de la sala de examen

Los colegios deben realizar los exámenes siguiendo unas reglas muy estrictas estipuladas en la publicación La realización de los exámenes del Programa del Diploma del IB (con la fecha de la convocatoria) disponible en el Centro de recursos para los programas > Implementación > Procesos y procedimientos de evaluación. Estas reglas cubren todos los aspectos de la organización de exámenes, desde la recepción del material de examen, pasando por la celebración de los exámenes, hasta el envío de las respuestas de los alumnos a un centro de escaneado.

El IB da una gran importancia a mantener la seguridad de los cuestionarios de examen y una conducta adecuada en los exámenes, ya que si los colegios u otras partes implicadas tienen la impresión de que ha habido algún caso de conducta improcedente, el valor de los resultados de los alumnos disminuirá considerablemente. Con el fin de mantener la confianza general, el IB investiga minuciosamente todos los avisos que denuncian casos de mala administración. Además, el personal o los asesores del IB realizan inspecciones aleatorias en los colegios durante las convocatorias de exámenes para comprobar la seguridad de los exámenes y asegurarse de que se realicen conforme a las normas de la organización. A pesar de todo, estos son controles esporádicos de los procesos, y el IB otorga una gran responsabilidad a los directores de colegio y a los coordinadores de los programas para evitar situaciones de mala administración, ya que estos profesionales son quienes están en la mejor posición para garantizar que las normas de calidad se mantengan en sus colegios día a día. Además, se espera que la cultura escolar del



colegio aliente a sus docentes y a sus alumnos a adoptar las mejores prácticas y la mayor integridad posibles.

Una de las peores situaciones posibles para los alumnos es realizar los exámenes y que el colegio o los servicios postales los pierdan. Si esto sucede, para intentar mitigar los perjuicios, el IB puede hacer una estimación de la puntuación que se les concede a los alumnos, aunque esto solo es posible si hay otras pruebas (cuestionarios de examen) en las que basarse. Para asegurase de que así sea, el IB requiere que los distintos cuestionarios de examen con las respuestas de los alumnos se envíen al centro de escaneado en días distintos, ya que así se reduce la posibilidad de que todos se pierdan por el camino.

La introducción de la evaluación en pantalla disminuye aún más este riesgo, ya que se dispondrá de copias del trabajo del alumno 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 de los alumnos, 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 de los alumnos, esto supone una distribución más uniforme de la carga de trabajo durante las tres semanas del calendario de exámenes, y les permite recuperarse durante la noche en caso de que crean que no han dado todo lo que pueden de sí mismos en un examen concreto.

#### Principios del diseño 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.

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

- Dada la gran cantidad de países en los que se ofrecen los programas del IB, actualmente no es posible tener en cuenta los días festivos públicos, nacionales o escolares, ni tampoco las festividades religiosas.
- 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 no es posible que así sea.
- Se tendrán 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 asegurarse de que la coincidencia de asignaturas en el calendario afecte a la menor cantidad de alumnos posible en todo el mundo.
- No se examinará a los alumnos 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 en todos los componentes de los cursos en los que hay varias pruebas de examen, dichas pruebas se programarán para realizarse en dos días como mínimo.

- Los alumnos deben 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 alumnos (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 tres semanas 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 alumnos.

# La 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 en la sección "Equidad para todos: cómo satisfacer las necesidades de los alumnos".

El IB cree que todos los alumnos deben tener la posibilidad de demostrar sus capacidades en condiciones de examen lo más equitativas posible. Las condiciones normales de evaluación pueden representar una desventaja para aquellos alumnos 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 cuestionarios de examen.

Se denomina circunstancia adversa a cualquier acontecimiento ajeno a la voluntad del alumno que ocurre durante la convocatoria de exámenes y que puede afectar negativamente en 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. 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. En esas situaciones, el colegio debe contactar con el IB y proporcionar detalles sobre las circunstancias para que el Comité de la evaluación final decida si las debe tener en cuenta al determinar la calificación de un alumno.

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 realizarían el examen bastante antes o después que otros, lo cual fácilmente podría conducir a alguna conducta improcedente. 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 de los alumnos que realicen el examen sería demasiado grande.

Existen solo tres circunstancias en las que el IB permitirá a un alumno realizar uno o varios exámenes a una hora o en una fecha diferente de la programada para los exámenes del IB:

- Coincidencia entre varios exámenes del IB programados para la misma fecha y la misma hora
- Coincidencia entre las fechas de los exámenes del IB y las de otras organizaciones educativas, incluidas las pruebas de ingreso a la universidad



#### Situaciones de emergencia

El cambio de hora o fecha de un examen se autorizará únicamente si el coordinador puede garantizar la seguridad del examen. Si se autoriza un cambio en el calendario para que el examen se realice el mismo día, antes o después de la hora programada, el coordinador debe asegurarse de que el alumno o los alumnos correspondientes son supervisados en todo momento entre la hora programada y la nueva. Esta medida sirve para garantizar que no haya comunicación con ningún alumno que ya haya realizado el mismo examen.

En los Procedimientos de evaluación del programa correspondiente se explica detalladamente cómo solicitar el cambio de fecha de un examen.

# Fijación del estándar: cómo preparar a los examinadores para la corrección

Los propósitos clave de fijar el estándar son:

- Que el examinador principal determine el estándar para la evaluación (a partir de consultas con otros examinadores) en función del trabajo de los alumnos
- Poner a prueba y perfeccionar el esquema de calificación para asegurarse de que permite recompensar de manera justa el esfuerzo del alumno
- Elaborar exámenes corregidos definitivos (de práctica, de aptitud y de control)
- Transmitir lo aprendido a todos los examinadores
- Confirmar que se entiende el estándar, para lo cual los examinadores pasan el proceso de corrección de respuestas de aptitud

La fijación del estándar es el período de tiempo que cubre varias partes del proceso de evaluación. El propósito de esta etapa es preparar a los examinadores para que su corrección sea eficaz y fiable.

Es importante que todos los examinadores concedan la misma puntuación a los trabajos que tengan la misma calidad porque, de lo contrario, el examinador que corrija el trabajo influirá en la calificación final que reciba el alumno. El examinador principal de cada componente fija el estándar, y el resto de examinadores deben respetarlo y regirse por él.

El proceso de fijar el mismo estándar para todos los examinadores se denomina estandarización.

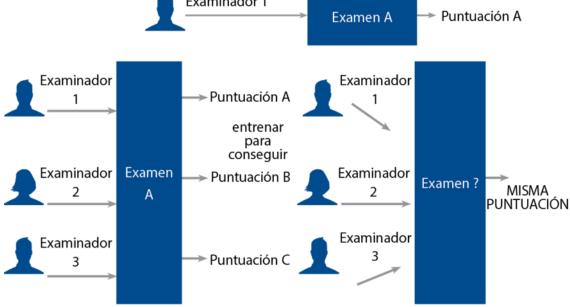
Figura 44

Todos los examinadores tienen que corregir los exámenes según el estándar del examinador principal.

Examinador 1

Franco A

Puntuación A



El primer paso del proceso de estandarización consiste en seleccionar a examinadores con los conocimientos y las habilidades necesarios. A continuación, se les ofrece una capacitación general para que



entiendan el proceso y los principios. Para obtener más información sobre la selección y capacitación de examinadores, consulte el sitio web del IB.

### Reunión de estandarización

Se conoce como reunión de estandarización a la discusión para fijar el estándar en la que participa el equipo de examinadores supervisores, dirigido por el examinador principal. Esta reunión puede ser presencial, pero la mayoría de las veces se hace virtualmente combinando diversos foros de debate del IB y videoconferencias para evitar que los examinadores supervisores, que están repartidos por todo el mundo, tengan que volar a Cardiff (Reino Unido). Los equipos de examinadores supervisores solamente tendrán que viajar a una reunión presencial si hay algún motivo concreto para hacerlo, como que sea la primera convocatoria de exámenes de un curso nuevo o que haya un proceso de estandarización especialmente complejo.

Después de que los alumnos realizan los exámenes, los examinadores supervisores revisan ejemplos del trabajo de los alumnos para comprobar que el esquema de calificación cubra adecuadamente la variedad de respuestas que pueden encontrarse los examinadores. Aunque el esquema de calificación no puede cubrir todas las respuestas posibles, el examinador principal y el equipo de examinadores supervisores deben asegurarse de que dicho esquema explique claramente el estándar requerido para cada puntuación.

Es importante que el esquema de calificación o las notas para la corrección no sean tan detallados que pierdan su eficacia porque los examinadores queden confundidos o pasen por alto determinados elementos fundamentales. Tampoco se debe caer en el extremo contrario: el esquema de calificación debe incluir suficiente información para garantizar que todos los examinadores concedan la misma puntuación al mismo tipo de respuesta. A título indicativo, si un comentario concreto es pertinente para menos del 10 % de los alumnos, es poco probable que sea una orientación útil.

De la reunión de estandarización se obtiene un resultado clave: una serie de **exámenes corregidos definitivos**. Se trata de ejemplos del trabajo de los alumnos que ha corregido el examinador principal y que se utilizan posteriormente para instruir y poner a prueba a otros examinadores en un proceso que se denomina *modelo de calidad*. Es muy importante que estos exámenes corregidos definitivos sean correctos, porque se utilizan para tomar decisiones sobre la calidad del trabajo de otros examinadores. Hay que considerar tres usos para los exámenes corregidos definitivos:

- 1. Exámenes de práctica
- 2. Exámenes de aptitud
- 3. Exámenes de control

## Exámenes de práctica

- El propósito de los exámenes de práctica es ayudar a los examinadores a aprender el estándar de corrección.
- Los ejemplos deben mostrar cómo corregir trabajos típicos de los alumnos o situaciones habituales en las que los examinadores no entienden correctamente el esquema de calificación.

Los examinadores utilizan estos ejemplos corregidos definitivos del trabajo de los alumnos para comprobar si han comprendido el esquema de calificación y las notas para la corrección. Una vez revisados estos exámenes, los examinadores deberían confiar en que son capaces de corregir de manera correcta.

Al seleccionar los exámenes de práctica, los examinadores principales piensan en:

- Ayudar a los examinadores a prepararse para corregir, de modo que no incluirán exámenes que representen una situación con la que no es probable que vuelvan a encontrarse
- Cómo explicar, mediante el uso de comentarios informativos, por qué se han otorgado las puntuaciones
- Si resulta útil y razonable excluir algunas partes del examen del alumno o indicar en una nota que no son de utilidad para el aprendizaje de los examinadores

Un buen conjunto de exámenes de práctica debe incluir (por orden de prioridad):

- 1. Ejemplos que muestren cómo corregir trabajos típicos de los alumnos
- 2. Situaciones habituales en las que los examinadores no entienden correctamente el esquema de calificación
- 3. Un buen rango de puntuaciones, de bajas a altas, para ayudar a los examinadores a determinar qué notas asignar en cada caso y qué se entiende por una buena (o mala) respuesta
- 4. Ejemplos de excepciones y respuestas complejas, y cómo trabajar con ellas

## Exámenes de aptitud

- El propósito de los exámenes de aptitud es demostrar o probar que los examinadores corrigen de acuerdo con el estándar adecuado.
- Dichos exámenes comprenden ejemplos de las respuestas más frecuentes de los alumnos en todo el rango de puntuaciones.

Antes de que se les autorice corregir el trabajo de los alumnos en condiciones reales, los examinadores deben demostrar que son capaces de corregir siguiendo el estándar establecido. Para ello, se les pone a prueba y se les pide que corrijan cinco ejemplos de trabajos de alumnos que ya ha corregido el examinador principal, por lo que se sabe de antemano qué puntuación se les debe otorgar.

Si no conceden las mismas puntuaciones que el examinador principal, los examinadores reciben comentarios sobre cuál era la puntuación correcta y por qué. Tras reflexionar sobre estos comentarios, se da una segunda oportunidad a los examinadores para que demuestren que han comprendido el estándar de corrección.

Un buen conjunto de exámenes de aptitud debe incluir (por orden de prioridad):

- 1. Ejemplos de las respuestas más frecuentes de los alumnos en todo el rango de puntuaciones (altas, medias y bajas)
- 2. Ejemplos de trabajos que requieran que el examinador haya comprendido las excepciones comunes al esquema de calificación o los errores habituales que cometen los alumnos

Al seleccionar los exámenes de aptitud, los examinadores principales no deben incluir exámenes que estén diseñados para "sorprender con la guardia baja" al examinador, sino que deben elegir exámenes difíciles de corregir pero que representen el tipo de dificultades que se espera que los examinadores resuelvan diariamente. Además, también deben incluirse ejemplos de casos importantes para tener la certeza de que los examinadores los gestionarán correctamente cuando se enfrenten a ellos durante el proceso real de corrección.

### Exámenes de control

- El propósito de los exámenes de control es demostrar o probar que los examinadores siguen corrigiendo de acuerdo con el estándar adecuado.
- Los exámenes de control incluyen errores comunes que cometen los examinadores y ejemplos de todo el rango de puntuaciones.

Aunque no se permite que ningún examinador empiece a corregir en condiciones reales los trabajos de los alumnos antes de que comprenda el estándar de corrección, es evidente que su manera de corregir puede variar con el paso del tiempo. Por ese motivo, la corrección de los examinadores se comprueba periódicamente mediante exámenes de control.

Estos son exámenes que ya ha corregido el examinador principal, por lo que se sabe de antemano qué puntuación merecen. Los exámenes de control tienen el mismo aspecto que el resto de los exámenes, por lo que el examinador no sabe que está corrigiendo un examen de control.



Si el examinador concede la misma puntuación que el examinador principal en el examen de control, podrá continuar corrigiendo. Sin embargo, si su corrección ya no sigue el estándar requerido, el IB interviene.

Inicialmente, se ofrecen comentarios y orientación para que el examinador tenga la oportunidad de ajustarse de nuevo al estándar de corrección. A continuación, el examinador puede volver a corregir, sabiendo que todos los alumnos recibirán una puntuación coherente, independientemente de quién corrija su trabajo.

Si el examinador no consigue volver a ajustarse al estándar de corrección adecuado, no se le permite que corrija ningún trabajo más.

En general, los examinadores deben corregir un examen de control por cada diez exámenes en condiciones reales. Sin embargo, no siempre se trata del décimo examen, con el fin de que los examinadores no sepan cuándo corrigen un examen de control. Si corresponde, se puede cambiar la frecuencia de los exámenes de control.

Un buen conjunto de exámenes de control debe incluir (por orden de prioridad):

- Trabajos de alumnos que representen errores comunes que cometen los examinadores
- Un buen rango de puntuaciones (altas, medias y bajas)

A diferencia de los exámenes de aptitud, los exámenes de control se asignan aleatoriamente, por lo que cada examinador los recibirá en un orden distinto.

Al seleccionar los exámenes de control, los examinadores principales son conscientes de que los examinadores están corrigiendo trabajos de los alumnos en condiciones reales, así que la atención se centra en garantizar que los alumnos reciban las puntuaciones correctas. Esto implica que resulta adecuado incluir exámenes que sean complejos o difíciles de corregir. El propósito de estos exámenes no es "sorprender con la guardia baja" al examinador, sino asegurarse de que corrige como se espera. También es razonable utilizar los exámenes de control para comprobar si los examinadores siguen todas las normas de corrección, y no solo para comprobar si la puntuación que otorgan es correcta. Esto significa que, en raras ocasiones, el examinador principal puede incluir un examen de control que deba remitirse al jefe de equipo por algún motivo (por ejemplo, si hay indicios de conducta improcedente o una respuesta totalmente inesperada).

Los jefes de equipo deben supervisar y apoyar a los examinadores a lo largo del proceso de corrección. Los jefes de equipo deben ponerse en contacto con los examinadores a los que se les impida corregir porque hayan corregido de manera incorrecta un examen de aptitud o de control, y deben proporcionarles comentarios y mentoría. El jefe de equipo y el examinador podrán ver el examen de control que no se haya corregido en consonancia con el estándar, y podrán consultar los comentarios y las puntuaciones del examinador principal. La función del jefe de equipo es explicar al examinador por qué las puntuaciones concedidas son incorrectas, y podrán discutir el esquema de calificación y su correcta aplicación. Si, a pesar de la capacitación adicional prestada por el jefe de equipo, los examinadores continúan aplicando incorrectamente el esquema de calificación, no se les permitirá continuar corrigiendo más.

# Márgenes de tolerancia

- El margen de tolerancia refleja las diferencias legítimas en las puntuaciones que distintos examinadores otorgan a un mismo trabajo. Imaginemos, por ejemplo, a dos profesores de un colegio que corrigen el mismo trabajo; ambos coinciden en que el trabajo es bueno, pero uno le daría una puntuación de 46 y el otro de 47.
- El IB tiene el compromiso de 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, se espera que el examinador conceda la misma puntuación que el examinador principal. En el caso de otros tipos de preguntas, sobre todo las que ponen a prueba la comprensión o el análisis, las puntuaciones concedidas dependerán en cierta medida del juicio del examinador. Dos examinadores expertos con el mismo

concepto de lo que es una buena respuesta concederían una puntuación parecida, si bien puede haber una diferencia de uno o dos puntos.

La idea de utilizar márgenes de tolerancia es reflejar estas diferencias de opinión legítimas al revisar el desempeño de los examinadores. Por ejemplo, si el examinador 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 que conceda una puntuación de entre 44 y 48 demuestra que corrige según el estándar establecido.

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 (es decir, el total) y también las diferencias en cada parte o criterio. El examinador demostrará que corrige según el estándar si sus resultados se encuentran dentro de los márgenes de tolerancia en ambos aspectos.

# Fijación satisfactoria del estándar

Si esta etapa del ciclo de evaluación se ha realizado de manera satisfactoria, debería haber:

- Coherencia en las correcciones
- Un grupo de jefes de equipo capaces de intervenir y ayudar a sus examinadores con casi cualquier consulta relacionada con la corrección de un componente
- Instrucciones, incluidos el esquema de calificación o las notas para la corrección, que expliquen claramente la fundamentación de las puntuaciones otorgadas
- Exámenes corregidos definitivos que sean adecuados al propósito (y que estén bien corregidos)



### Corrección

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

# Comprender la corrección como el resultado del juicio coherente de los examinadores

El propósito del esquema de calificación es ayudar a los examinadores y recordarles el estándar de corrección. El esquema de calificación no define en qué casos se deben conceder puntuaciones. El juicio del examinador principal es la definición de la puntuación "correcta", y los esquemas de calificación son el modo que tiene de explicar su pensamiento.

Este principio es muy importante, porque significa que si la opinión del examinador principal no concuerda con el esquema de calificación o con las notas para la corrección, entonces lo incorrecto es el esquema de calificación y se deberá enmendar.

Afortunadamente, hay muy pocos casos en los que esto suceda, pero son extremadamente graves, porque significa que habría que volver a informar a los demás examinadores que habían confiado en el esquema de calificación sobre cuál debería ser el estándar de corrección adecuado. Además, el IB tendría que volver a corregir todos los trabajos que pudieran haberse visto afectados por el error, a fin de que no hubiera ningún alumno con una puntuación incorrecta. Los examinadores principales y el equipo supervisor son muy cuidadosos al fijar el estándar, para garantizar que el esquema de calificación sea preciso e informativo y también para evitar este tipo de situación.

Los examinadores deben haber corregido satisfactoriamente los exámenes de aptitud y los de control antes de poder corregir los exámenes reales. Para poder corregir trabajos de los alumnos en situaciones reales, no basta con saber interpretar el esquema de calificación. Asimismo, los profesores y los padres no pueden dar por hecho que su propia interpretación del esquema de calificación sirve para demostrar que un examen se ha corregido erróneamente. En esos casos, se les invita a utilizar el servicio de consulta sobre los resultados.

#### "Puntuaciones definitivas"

Los términos *puntuaciones definitivas* o *examen corregido definitivo* implican que el examinador principal ha emitido un juicio formal sobre la puntuación que se debe conceder. Los exámenes corregidos definitivos se utilizan en el modelo de calidad.

# Puntuación y calificación son conceptos distintos

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

 Al corregir, se reconoce la labor que ha realizado un alumno y se le otorgan unas puntuaciones según un esquema de calificación o un marco similar. Esto indica en qué medida el alumno realizó correctamente la tarea de evaluación. La puntuación en sí misma no significa nada más.  A la hora de decidir la calificación, el examinador juzga la calidad del trabajo del alumno 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.

Figura 10 calidad (el sabor) de una comida aue se ha cocinado, mientras aue la

La puntuación sería la calidad (el sabor) de una comida que se ha cocinado, mientras que la calificación refleja la complejidad del plato que se quería cocinar.



Es posible que un alumno reciba una calificación alta al responder bien solamente una parte pequeña de una pregunta muy difícil, pero que no pueda conseguir la misma calificación si solo responde correctamente muchas preguntas triviales.

Tal como se indica en secciones posteriores, no es necesario hacer referencia a los logros del alumno 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 en comparación con sus compañeros.

En sus evaluaciones, el IB normalmente utiliza la puntuación como indicativo del desempeño general (modelo de compensación), y posteriormente estudia el desempeño de los alumnos con esa puntuación para determinar una puntuación límite (límite de calificación) en la que se concede una calificación concreta a los alumnos que superan una puntuación determinada. Este proceso se explica con más detalle en la sección "Las prácticas de evaluación del IB".

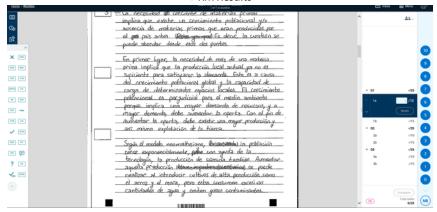
# Principios básicos de la corrección en pantalla

- La corrección en pantalla permite mejorar la calidad de la corrección de las evaluaciones.
- Permite que el IB supervise la corrección de los examinadores al instante.
- Elimina el tiempo necesario para enviar y devolver los exámenes con las respuestas de los alumnos, así como el riesgo de que se pierdan.
- Permite que el IB divida en varias partes los exámenes escritos de los alumnos y envíe distintas preguntas a correctores expertos (grupos de preguntas).
- Los examinadores pueden corregir materiales de audio y video y evaluaciones en pantalla, así como exámenes manuscritos.

En la corrección en pantalla se presentan en formato electrónico imágenes del trabajo de los alumnos a los examinadores, a quienes se les pide que corrijan el trabajo directamente en una computadora. Para ello, utilizan un *software* (herramienta de corrección) que les permite añadir signos de visto, puntuaciones y anotaciones en el examen electrónico, igual que harían en una versión en papel.



Figura 45 Captura de pantalla de RM Assessor, la herramienta de corrección electrónica que se utiliza actualmente en el IB. © RM Results



La computadora registra directamente las puntuaciones, por lo que se puede supervisar a los examinadores mientras corrigen. Asimismo, tener en formato electrónico los exámenes con las respuestas de los alumnos ofrece otras ventajas considerables.

En primer lugar, los exámenes de los alumnos pueden transmitirse rápidamente a los examinadores, con lo cual se elimina la espera provocada por el envío de la versión en papel. Además, se elimina también tener que adivinar cuántos exámenes podrá corregir un examinador, ya que ahora se le pueden ir proporcionando a medida que los necesite. Disponer de copias de los trabajos de los alumnos en el sistema es un beneficio considerable, ya que así se asegura que no se pierdan ni se dañen.

En segundo lugar, esto permite introducir exámenes de prueba (denominados *exámenes de control*) para poder comprobar si los examinadores respetan sistemáticamente el estándar a lo largo de su trabajo de corrección.

En tercer lugar, permite que el IB divida los exámenes con las respuestas de los alumnos en preguntas individuales (denominadas *grupos de preguntas*) y que se pueda pedir a los examinadores que se centren en corregir preguntas individuales en vez de exámenes completos.

Por último, la corrección en pantalla posibilita utilizar material de video y audio. Esto amplía considerablemente la variedad de trabajos de alumnos que se pueden enviar para demostrar sus conocimientos y habilidades como, por ejemplo, grabaciones de representaciones o de exámenes orales.

# Distintos tipos de esquemas de calificación

#### Esquemas de calificación analíticos

Los esquemas de calificación analíticos se preparan para aquellas preguntas de examen que se espera que los alumnos contesten con una variedad limitada de respuestas.

Estos esquemas de calificación pueden dar instrucciones específicas a los examinadores 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 suficiente información para poder corregir de manera coherente los principales tipos de enfoques que podría adoptar el alumno, así como los errores frecuentes que podría cometer. Inevitablemente, los examinadores 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.

En muchas ocasiones, no es posible predecir qué partes de una pregunta responderán bien o mal los alumnos. 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 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 cuando haya cometido un error en una parte de una pregunta.

Esto es particularmente problemático en asignaturas como Ciencias o Matemáticas, en las que el IB espera que los examinadores concedan puntuaciones a las respuestas parcialmente correctas. Sin la estructura propia de una pregunta que requiere una respuesta en profundidad, es posible que algunos alumnos no obtengan muchos puntos porque cometan un leve error al inicio de su respuesta, o porque un ligero problema de comprensión al leer la pregunta los lleve en una dirección equivocada. A menudo, los esquemas de calificación analíticos más elaborados son los de Matemáticas.

#### Criterios de evaluación

Si una tarea de evaluación es tan abierta que la variedad de posibles respuestas es demasiado amplia como para que los esquemas de calificación analíticos sean eficaces, entonces se aplican en su lugar criterios (de evaluación) descriptivos.

Por lo general, los criterios de evaluación no se refieren al contenido específico que se espera en la respuesta de un alumno. Sin embargo, en la medida de lo posible, sí que mencionan la necesidad de que los alumnos muestren determinados tipos de conocimientos del contenido. En realidad, los criterios se centran más en el tipo de desempeño que se espera que demuestren los alumnos, 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. Basándose en los descriptores, el examinador 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.

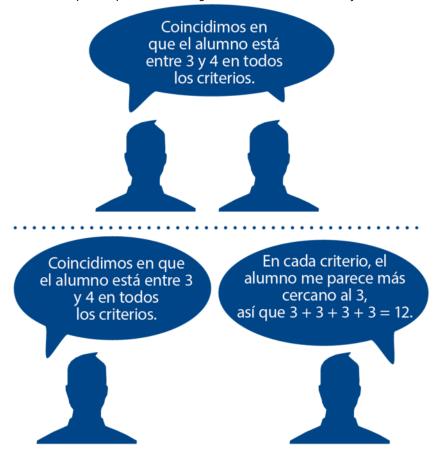
Aunque es perfectamente posible corregir un trabajo usando un solo criterio, lo más frecuente es emplear varios criterios. Una vez hecho esto, se suman las puntuaciones individuales (hay que tener en cuenta que los criterios que se consideran más importantes tienen una puntuación total más alta). En esos casos, es muy importante que los criterios sean independientes los unos de los otros, ya que no sería justo que un alumno recibiera puntos en distintas partes por el mismo aspecto de su respuesta.

Utilizar varios criterios plantea otro problema, y es que a menudo esto reduce la fiabilidad y la coherencia de la corrección. Esto se debe a que es más probable que haya respuestas en las que el examinador tenga que decidir si están en un nivel de logro o en otro y, aunque la diferencia sea de un solo punto, esta situación se puede magnificar si todas las diferencias son al alza o a la baja.



Figura 46

Muestra de cómo los criterios de evaluación pueden llevar a los examinadores a otorgar puntuaciones muy distintas aunque compartan su visión general de la calidad del trabajo del alumno.



Es habitual reutilizar los mismos criterios de evaluación cada año, aunque cambien las preguntas que se plantean a los alumnos. Esto se debe a que la naturaleza subyacente de lo que se evalúa no varía, y las preguntas nuevas normalmente tratan sobre detalles específicos que no figuran explícitamente en los criterios (a diferencia de lo que sucede en los esquemas de calificación analíticos). El examinador principal suele proporcionar notas para la corrección que sirven para explicar a los examinadores cómo aplicar los criterios de evaluación en cada pregunta, y es posible que también proporcione ejemplos de detalles pertinentes. En los casos en que los criterios de evaluación se utilicen para la evaluación interna, los profesores y los moderadores deben consultar el material de ayuda al profesor publicado, que ofrece varios ejemplos sobre cómo aplicar los criterios.

Es importante tener en cuenta que, aunque los descriptores de nivel de un criterio tienen una naturaleza jerárquica y a menudo se refieren a la jerarquía de habilidades cognitivas definida por Bloom *et al.* (1956), no existe ninguna relación directa entre la exigencia cognitiva y los niveles de los criterios. Los descriptores de nivel más bajos no abordan únicamente las habilidades cognitivas "más sencillas", y los descriptores de nivel más altos no están reservados únicamente para las habilidades cognitivas de "orden superior". Cada una de las áreas de habilidades cognitivas descritas por Bloom debe contar con una gama de niveles de logro.

#### 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 de los alumnos, cada descriptor de nivel de la banda de puntuación corresponde a un rango de puntuaciones diferentes.

Figura 47
Ejemplo de bandas de puntuación obtenido de la Guía de Historia del PD de 2015, prueba 1 de NM y NS

Puntuación	Descriptor de nivel
0	La respuesta no alcanza ninguno de los niveles especificados por los descriptores que figuran a continuación.
1–3	La respuesta no se centra en la pregunta.  Si bien se incluyen referencias a las fuentes, es probable que a este nivel las referencias consistan en descripciones del contenido de las fuentes y no se utilicen como pruebas para respaldar el análisis.  No se demuestran conocimientos propios o, si quedan demostrados, son incorrectos o carecen de pertinencia.
4–6	En su mayor parte, la respuesta se centra en la pregunta.  Se incluyen referencias a las fuentes, y esas referencias se utilizan como pruebas para respaldar el análisis.  Los conocimientos propios demostrados carecen de precisión o pertinencia. No se intenta o apenas se intenta sintetizar los conocimientos propios y la información proveniente de las fuentes.
7–9	La respuesta se centra en la pregunta.  Se incluyen referencias claras a las fuentes, y esas referencias se utilizan eficazmente como pruebas para respaldar el análisis.  Se utilizan conocimientos propios precisos y pertinentes. Se sintetizan eficazmente los conocimientos propios y la información proveniente de las fuentes.

En general, los descriptores suelen ser bastante largos porque cubren una variedad de cualidades que pueden aparecer en el trabajo de los alumnos, 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. Además, los examinadores 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 se ajusta a dicho descriptor. Por ejemplo, un nivel de una banda de puntuación puede cubrir el rango de 6 a 10 puntos. El examinador concederá al trabajo del alumno la puntuación de ese rango que mejor se ajuste al descriptor de nivel pertinente.

Hay investigaciones que muestran que cuando se utilizan tanto criterios de evaluación como una corrección holística (bandas de puntuación) para evaluar ensayos que se prestan a ambos métodos, hay pocas diferencias entre dichos métodos en lo referido a la fiabilidad de la corrección (Wood, 1991).

# Apoyo adicional para los examinadores

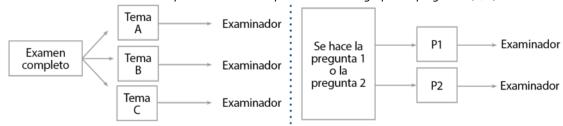
Al igual que todos los actores del aprendizaje, los examinadores se benefician de contar con distintas maneras de asimilar información.

Además de los esquemas de calificación y los criterios de evaluación disponibles en formato escrito, el IB cada vez utiliza más videos para explicar con medios audiovisuales el estándar de corrección requerido. Por otra parte, también se anima a los examinadores a que hagan preguntas al equipo de examinadores supervisores, tanto por teléfono, como por mensajes a través del *software* de corrección electrónica o correo electrónico.

Esta comunicación está integrada en el modelo de calidad del IB. Cuando, al corregir un examen de aptitud o de control, un examinador no se ajusta al estándar establecido, su jefe de equipo obligatoriamente debe proporcionarle los comentarios pertinentes. Si la corrección es conforme al estándar establecido, el examinador tiene igualmente la oportunidad de recibir comentarios.

# Grupos de preguntas

Figura 48
Las dos formas en que los exámenes se pueden dividir en grupos de preguntas (QIG)



Las dos formas en que los exámenes se pueden dividir en grupos de preguntas (QIG)

La razón de ser de los grupos de preguntas (QIG, por sus siglas en inglés) es que para los examinadores es más fácil corregir la misma pregunta varias veces seguidas que corregir todas las preguntas de un mismo examen y luego empezar de nuevo con la primera pregunta para el siguiente alumno.

Hay dos formas de dividir los exámenes en QIG. La primera consiste en dividir los exámenes según la pregunta o el área temática. En este caso, es importante que las partes de las preguntas que guarden relación entre sí se incluyan en el mismo QIG. Por ejemplo, si la parte "a" lleva a la parte "b", ambas deben formar parte del mismo QIG. Así, los examinadores pueden otorgar puntos por coherencia o arrastre de error, y también conceder puntos a los alumnos que aportan en la primera parte de su respuesta elementos pertinentes para la segunda parte.

Sin embargo, cuantos menos detalles del esquema de calificación tenga que recordar el examinador, más coherente será en su corrección, por lo cual es preferible que haya varios QIG pequeños a que haya uno grande. Este aspecto se tiene en cuenta cuando se decide cómo dividir los cuestionarios de examen.

La segunda manera de dividir un examen en QIG tiene lugar cuando el alumno puede elegir entre varias preguntas que responder. En este caso, cada una de las preguntas opcionales se convierte en un QIG, que puede conformar la totalidad del examen del alumno. Por ejemplo, imaginemos que en Literatura un alumno puede elegir una pregunta sobre poesía u otra sobre prosa para redactar un ensayo. Todas las preguntas sobre poesía conformarían el QIG 1, y un examinador experto en poesía sería el encargado de corregir todas estas respuestas. Por su parte, todas las preguntas sobre prosa conformarían el QIG 2, y un examinador experto en prosa sería el encargado de corregir esas respuestas.

Cada QIG contará con sus propios exámenes de prueba, de aptitud y de control, y el examinador deberá demostrar que es capaz de corregir según el estándar establecido para su QIG. Aunque pueda parecer complicado, esto significa que un examinador 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.

#### Modelo de calidad

En la actualidad, el IB corrige electrónicamente casi todo el trabajo de evaluación externa. El desempeño de los examinadores se comprueba mediante un sistema que requiere que estos y el examinador principal corrijan los mismos exámenes de los alumnos, y que a continuación compara las puntuaciones de los

examinadores con las del examinador principal en el componente en cuestión. El proceso funciona de la siguiente forma:

- El examinador principal y un grupo reducido de examinadores supervisores corrigen varios exámenes durante la semana después de que los alumnos hayan realizado el examen, como parte del proceso de estandarización.
- 2. Las puntuaciones definitivas que otorga el equipo de examinadores supervisores se registran en el sistema junto con sus comentarios o anotaciones.
- 3. El IB, con el asesoramiento del equipo de examinadores supervisores, determina la diferencia aceptable entre la puntuación definitiva (la que concede el examinador principal) y la puntuación que otorgan los examinadores en estos exámenes. Esta diferencia permitida se denomina margen de tolerancia.
- 4. A continuación, los comentarios y las puntuaciones se ocultan, y los exámenes vuelven a colocarse con el resto para que los examinadores los corrijan.
- 5. Los exámenes corregidos definitivos se usan para las correcciones de práctica, para el proceso de estandarización y para el sistema de control.

La idea en la que se basa el modelo de calidad es que los examinadores pueden aprender durante todo el proceso de corrección: cuando corrigen los exámenes de control fuera del margen de tolerancia, se les muestra inmediatamente las puntuaciones definitivas para que entiendan que no han aplicado correctamente el esquema de calificación e identifiquen qué deben mejorar. Además, a los examinadores también se les muestran periódicamente los exámenes de control que han corregido dentro de los márgenes de tolerancia, con el fin de que puedan acercarse lo máximo posible a la puntuación definitiva.

Los examinadores deben, por tanto, considerar los exámenes de control como oportunidades de desarrollo profesional. La mayoría de los examinadores necesitarán comentarios sobre los exámenes de control en algún momento del proceso de corrección, y esos comentarios deben valorarse en vez de ser motivo de preocupación.

## Exámenes de práctica

- El propósito de los exámenes de práctica es ayudar a los examinadores a aprender el estándar de corrección.
- Los ejemplos deben mostrar cómo corregir trabajos típicos de los alumnos o situaciones habituales en las que los examinadores no entienden correctamente el esquema de calificación.

Los examinadores utilizan estos ejemplos corregidos definitivos del trabajo de los alumnos para comprobar si han comprendido el esquema de calificación y las notas para la corrección. Una vez revisados estos exámenes, los examinadores deberían confiar en que son capaces de corregir de manera correcta.

Al seleccionar los exámenes de práctica, los examinadores principales piensan en:

- Ayudar a los examinadores a prepararse para corregir, de modo que no incluirán exámenes que representen una situación con la que no es probable que vuelvan a encontrarse
- Cómo explicar, mediante el uso de comentarios informativos, por qué se han otorgado las puntuaciones
- Si resulta útil y razonable excluir algunas partes del examen del alumno o indicar en una nota que no son de utilidad para el aprendizaje de los examinadores

Un buen conjunto de exámenes de práctica debe incluir (por orden de prioridad):

- 1. Ejemplos que muestren cómo corregir trabajos típicos de los alumnos
- 2. Situaciones habituales en las que los examinadores no entienden correctamente el esquema de calificación
- 3. Un buen rango de puntuaciones, de bajas a altas, para ayudar a los examinadores a determinar qué notas asignar en cada caso y qué se entiende por una buena (o mala) respuesta



Ejemplos de excepciones y respuestas complejas, y cómo trabajar con ellas

# Exámenes de aptitud

- El propósito de los exámenes de aptitud es demostrar o probar que los examinadores corrigen de acuerdo con el estándar adecuado.
- Dichos exámenes comprenden ejemplos de las respuestas más frecuentes de los alumnos en todo el rango de puntuaciones.

Antes de que se les autorice corregir el trabajo de los alumnos en condiciones reales, los examinadores deben demostrar que son capaces de corregir siguiendo el estándar establecido. Para ello, se les pone a prueba y se les pide que corrijan cinco ejemplos de trabajos de alumnos que ya ha corregido el examinador principal, por lo que se sabe de antemano qué puntuación se les debe otorgar.

Si no conceden las mismas puntuaciones que el examinador principal, los examinadores reciben comentarios sobre cuál era la puntuación correcta y por qué. Tras reflexionar sobre estos comentarios, se da una segunda oportunidad a los examinadores para que demuestren que han comprendido el estándar de corrección.

Un buen conjunto de exámenes de aptitud debe incluir (por orden de prioridad):

- Ejemplos de las respuestas más frecuentes de los alumnos en todo el rango de puntuaciones (altas, medias y bajas)
- Ejemplos de trabajos que requieran que el examinador haya comprendido las excepciones comunes al esquema de calificación o los errores habituales que cometen los alumnos

Al seleccionar los exámenes de aptitud, los examinadores principales no deben incluir exámenes que estén diseñados para "sorprender con la guardia baja" al examinador, sino que deben elegir exámenes difíciles de corregir pero que representen el tipo de dificultades que se espera que los examinadores resuelvan diariamente. Además, también deben incluirse ejemplos de casos importantes para tener la certeza de que los examinadores los gestionarán correctamente cuando se enfrenten a ellos durante el proceso real de corrección.

### Exámenes de control

- El propósito de los exámenes de control es demostrar o probar que los examinadores siguen corrigiendo de acuerdo con el estándar adecuado.
- Los exámenes de control incluyen errores comunes que cometen los examinadores y ejemplos de todo el rango de puntuaciones.

Aunque no se permite que ningún examinador empiece a corregir en condiciones reales los trabajos de los alumnos antes de que comprenda el estándar de corrección, es evidente que su manera de corregir puede variar con el paso del tiempo. Por ese motivo, la corrección de los examinadores se comprueba periódicamente mediante exámenes de control.

Estos son exámenes que ya ha corregido el examinador principal, por lo que se sabe de antemano qué puntuación merecen. Los exámenes de control tienen el mismo aspecto que el resto de los exámenes, por lo que el examinador no sabe que está corrigiendo un examen de control.

Si el examinador concede la misma puntuación que el examinador principal en el examen de control, podrá continuar corrigiendo. Sin embargo, si su corrección ya no sigue el estándar requerido, el IB interviene.

Inicialmente, se ofrecen comentarios y orientación para que el examinador tenga la oportunidad de ajustarse de nuevo al estándar de corrección. A continuación, el examinador puede volver a corregir, sabiendo que todos los alumnos recibirán una puntuación coherente, independientemente de quién corrija su trabajo.

Si el examinador no consigue volver a ajustarse al estándar de corrección adecuado, no se le permite que corrija ningún trabajo más.

En general, los examinadores deben corregir un examen de control por cada diez exámenes en condiciones reales. Sin embargo, no siempre se trata del décimo examen, con el fin de que los examinadores no sepan cuándo corrigen un examen de control. Si corresponde, se puede cambiar la frecuencia de los exámenes de control.

Un buen conjunto de exámenes de control debe incluir (por orden de prioridad):

- 1. Trabajos de alumnos que representen errores comunes que cometen los examinadores
- 2. Un buen rango de puntuaciones (altas, medias y bajas)

A diferencia de los exámenes de aptitud, los exámenes de control se asignan aleatoriamente, por lo que cada examinador los recibirá en un orden distinto.

Al seleccionar los exámenes de control, los examinadores principales son conscientes de que los examinadores están corrigiendo trabajos de los alumnos en condiciones reales, así que la atención se centra en garantizar que los alumnos reciban las puntuaciones correctas. Esto implica que resulta adecuado incluir exámenes que sean complejos o difíciles de corregir. El propósito de estos exámenes no es "sorprender con la guardia baja" al examinador, sino asegurarse de que corrige como se espera. También es razonable utilizar los exámenes de control para comprobar si los examinadores siguen todas las normas de corrección, y no solo para comprobar si la puntuación que otorgan es correcta. Esto significa que, en raras ocasiones, el examinador principal puede incluir un examen de control que deba remitirse al jefe de equipo por algún motivo (por ejemplo, si hay indicios de conducta improcedente o una respuesta totalmente inesperada).

Los jefes de equipo deben supervisar y apoyar a los examinadores a lo largo del proceso de corrección. Los jefes de equipo deben ponerse en contacto con los examinadores a los que se les impida corregir porque hayan corregido de manera incorrecta un examen de aptitud o de control, y deben proporcionarles comentarios y mentoría. El jefe de equipo y el examinador podrán ver el examen de control que no se haya corregido en consonancia con el estándar, y podrán consultar los comentarios y las puntuaciones del examinador principal. La función del jefe de equipo es explicar al examinador por qué las puntuaciones concedidas son incorrectas, y podrán discutir el esquema de calificación y su correcta aplicación. Si, a pesar de la capacitación adicional prestada por el jefe de equipo, los examinadores continúan aplicando incorrectamente el esquema de calificación, no se les permitirá continuar corrigiendo más.

## Márgenes de tolerancia

- El margen de tolerancia refleja las diferencias legítimas en las puntuaciones que distintos examinadores otorgan a un mismo trabajo. Imaginemos, por ejemplo, a dos profesores de un colegio que corrigen el mismo trabajo; ambos coinciden en que el trabajo es bueno, pero uno le daría una puntuación de 46 y el otro de 47.
- El IB tiene el compromiso de 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, se espera que el examinador conceda la misma puntuación que el examinador principal. En el caso de otros tipos de preguntas, sobre todo las que ponen a prueba la comprensión o el análisis, las puntuaciones concedidas dependerán en cierta medida del juicio del examinador. Dos examinadores expertos con el mismo concepto de lo que es una buena respuesta concederían una puntuación parecida, si bien puede haber una diferencia de uno o dos puntos.

La idea de utilizar márgenes de tolerancia es reflejar estas diferencias de opinión legítimas al revisar el desempeño de los examinadores. Por ejemplo, si el examinador 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 que conceda una puntuación de entre 44 y 48 demuestra que corrige según el estándar establecido.



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 (es decir, el total) y también las diferencias en cada parte o criterio. El examinador demostrará que corrige según el estándar si sus resultados se encuentran dentro de los márgenes de tolerancia en ambos aspectos.

# Respuestas complicadas o inusuales

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

Asimismo, el examinador principal será el encargado de corregir los exámenes que tengan respuestas inusuales y los exámenes de los alumnos que utilicen cuestionarios modificados. El examinador principal deberá velar por que el estándar de corrección en estos exámenes sea equivalente al que se aplica en todos los demás.

Esta corrección se realiza después de haberse determinado los límites de calificación, con el fin de que los examinadores supervisores tengan una comprensión común de las calificaciones en esa convocatoria de exámenes en concreto. Aunque al alumno se le concede una calificación, en estos casos concretos se pide a los examinadores que tengan en cuenta los límites de calificación para asegurarse de que los estándares de corrección establecidos específicamente para estos alumnos se adaptan a los límites de calificación establecidos para el grupo principal.

# Vínculos con los colegios

El IB parte del principio de que los examinadores no deben corregir el trabajo de sus propios alumnos.

Para ello, se pide a todos los examinadores que informen al IB de cualquier conexión que puedan tener con alumnos y colegios, a fin de que no se les envíe trabajo de estos para corregir. Asimismo, la división de evaluación del IB dispone de una política sobre conflictos de intereses que se aplica a todas las personas que trabajan para el IB susceptibles de tener vínculos con colegios.

Si sospecha que algún examinador o algún miembro del personal del IB no ha declarado sus vínculos con algún alumno o algún colegio, póngase en contacto con complaints@ibo.org o con "El IB responde".

#### Comentarios de los examinadores

No es posible ocuparse de dos cosas a la vez. Es como las calabazas en el agua: una sale a la superficie mientras se intenta mantener la otra hundida.

Proverbio chino.

El propósito de la evaluación sumativa del IB es medir el desempeño de los alumnos. Por tanto, se pide a los examinadores que se centren únicamente en corregir el trabajo de los alumnos según el estándar establecido. Los examinadores solo deben incluir comentarios si esto les ayuda en su labor de corrección.

Para poder ofrecer comentarios formativos, ya sean para los alumnos o para los profesores, el examinador debe determinar la puntuación correcta que corresponda al trabajo del alumno y, a continuación, intentar explicar cómo se podría mejorar dicho trabajo. 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 de su cometido principal, que es corregir según un estándar coherente. En términos sencillos, el IB quiere que los examinadores realicen una tarea (corregir) con la máxima calidad, y no dos tareas (corregir y proporcionar comentarios) con una calidad menor.

El examinador solamente puede ofrecer su valoración a partir del trabajo que tiene a su disposición, mientras que los profesores con experiencia pueden utilizar una amplia variedad de información para decidir qué comentarios aportar a un alumno. Esto significa que los comentarios del examinador tendrán una perspectiva más limitada que los del profesor.

Por todos estos motivos, el IB deja muy claro a los examinadores que deben limitarse a corregir el trabajo de los alumnos siguiendo el estándar establecido, y que no deben añadir comentarios para el alumno ni para el profesor.

Los examinadores 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 puntos sea más transparente para los colegios.

El informe de categoría 1, que los colegios pueden solicitar en el contexto de una consulta sobre los resultados, constituye la única excepción a este principio de proporcionar comentarios solo para explicar la corrección. En ese caso, un examinador supervisor responderá las inquietudes concretas que haya señalado el colegio al solicitar la consulta sobre los resultados, y aportará comentarios más detallados que los que normalmente se esperan de los examinadores.

#### 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 los alumnos pueden compensar un mal desempeño en un componente o pregunta si obtienen un buen resultado en otro.
- La calificación final que un alumno 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.



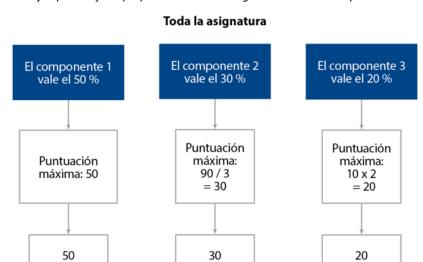


Figura 59
Ejemplo de ajuste proporcional en una asignatura con tres componentes

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 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.

Hay que recordar que no es necesario que las puntuaciones individuales del alumno coincidan con el porcentaje del total de la evaluación del componente. El IB es consciente de que no todos los alumnos tienen los mismos puntos fuertes, por lo cual se utilizan diversos instrumentos de evaluación. Es importante comparar los porcentajes del total de la evaluación reales de los resultados del grupo entero de alumnos con los porcentajes que se establecieron al diseñar la asignatura. Si, en una convocatoria de exámenes, la contribución de un componente es mucho mayor de lo que se esperaba, eso puede ser señal de que el examen era especialmente fácil y que se deben utilizar otras pruebas del desempeño de los alumnos para poder conceder las calificaciones finales. Por otro lado, si la contribución de un componente es habitualmente mucho mayor de lo que se esperaba, entonces se deberá revisar el diseño, ya que eso indicaría que la evaluación no funciona tal como se había previsto.

Es posible que este enfoque no refleje los métodos más sofisticados de ponderación, de combinación (totalización) y de ajuste proporcional que describió, por ejemplo, Wood (1991, capítulo 10), pero emplea los principios sólidos de la evaluación basada en criterios y respalda los principios de evaluación del IB.

El uso de la totalización para obtener la puntuación final implica que un alumno puede compensar un mal resultado en un componente con un buen resultado en otro, ya que lo único que afecta a la calificación final es la puntuación total. Este concepto se denomina *modelo de compensación*, en contraposición al *modelo de dominio*, en el que el alumno tiene que demostrar el nivel requerido en todos los componentes para poder obtener una calificación final concreta.

Merece la pena destacar que la calificación final que un alumno obtiene en una asignatura se determina a partir de la totalización de las puntuaciones logradas en cada componente, y no de calificaciones por componente. Como cada calificación en un componente representa un rango de puntuaciones, es posible

que dos alumnos con las mismas calificaciones en los componentes reciban distintas calificaciones en la asignatura.



### Moderación

- La moderación consiste en comprobar el estándar de corrección de los profesores, y no en volver a corregir el trabajo del alumno.
- Que una moderación esté bien hecha significa que los alumnos recibirían la misma puntuación de evaluación interna aunque hubieran estudiado en otro colegio en la otra punta del mundo. Esto se denomina estándar global.
- Entre las pruebas que se utilizan en la moderación figuran las explicaciones de los profesores acerca de por qué han otorgado sus puntuaciones, y no solo la calidad del trabajo del alumno.

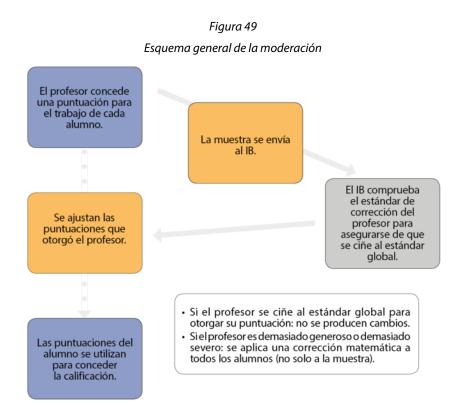
Gracias a la toma dinámica de muestras, que es un nuevo enfoque de la moderación, el IB se asegura de que los moderadores siguen el estándar del examinador principal, de modo que la moderación solo requiere una etapa.

# En qué consiste la moderación

Muchas veces no es posible evaluar un aspecto concreto de un alumno mediante un examen formal y con limitación de tiempo. En dichos casos, la forma más válida de proceder es pedir al profesor 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 tengan distintas interpretaciones del estándar de corrección. En ese sentido, sería posible que dos profesores de colegios distintos otorgaran puntuaciones distintas al mismo trabajo. El IB sigue un proceso integral de capacitación y puesta a prueba de los examinadores, con la intención de que todos compartan una comprensión común del estándar que se requiere. Sin embargo, esto no es posible para todos los profesores.

En lugar de ello, el IB confía en la capacidad de los profesores para corregir todos los trabajos de sus alumnos con coherencia, por lo que, si es necesario, solo hay que aplicar un factor aritmético para que las puntuaciones que concedan estén en consonancia con el estándar global. Para hacer esto, se pide una muestra de las correcciones a los profesores y se compara con el estándar del examinador principal. A partir de los datos que se desprenden de esta comparación, cuando es necesario se calcula una fórmula matemática que permite ajustar las puntuaciones que ha otorgado cada profesor.



Este diagrama explica la manera en que el examinador modera la muestra. 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 otorgó en ese componente a todos los alumnos del colegio. Si las puntuaciones que otorga el profesor 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 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.

#### Información importante que hay que recordar

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

Por razones prácticas, el IB modera a los colegios y no a profesores individuales. Por ese motivo, es muy importante que todos los profesores de un colegio se aseguren de corregir según el mismo estándar, de modo que el IB sea justo cuando aplique un factor de moderación a las puntuaciones de todos los profesores.



#### Figura 50

Aspectos que debe tener en cuenta un coordinador cuando revise las puntuaciones de la evaluación interna que se moderarán

En el colegio, ¿**todos** los profesores encargados de otorgar las puntuaciones de la evaluación interna corrigen según el mismo estándar?

¿Los profesores corrigen de manera coherente a todos los alumnos según el mismo estándar?

¿Los profesores explican claramente por qué conceden una puntuación concreta?

\* Los profesores encargados de otorgar las puntuaciones de la evaluación interna



Para obtener más información sobre los aspectos técnicos de cómo se calcula el factor de moderación, consulte el "Anexo 1: Moderación de la evaluación interna".

### Selección de los trabajos de los alumnos

- Todos los alumnos deben recibir una puntuación justa, por lo que el trabajo de cualquier alumno debe poder ilustrar el estándar del profesor y servir de ejemplo para la moderación.
- Es posible que distintos profesores tengan expectativas diferentes de lo que representa un muy buen trabajo o un muy mal trabajo, por lo que es importante que el material de muestra cubra todo el rango de puntuaciones.
- 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.

Con el objetivo de garantizar la transparencia y evitar cualquier impresión de conducta improcedente, 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 (véase a continuación) para garantizar que el factor de moderación sea lo más fiable posible y, dentro de las limitaciones de estos principios, que los alumnos se seleccionen 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 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 alumnos que se pida incluir en la muestra sea diferente en función del colegio si el IB necesita más ejemplos de la corrección de los profesores para determinar un factor de moderación sólido. Por ese motivo, a veces se piden muestras adicionales a algunos colegios. Puesto que es posible que el IB deba solicitar trabajos de muestra adicionales durante el proceso de moderación, el trabajo de todos los alumnos debe estar disponible hasta que se publiquen los resultados.

El segundo principio es que hay que estar seguro de que el factor de moderación es justo para todos los alumnos en todas las puntuaciones posibles. La experiencia demuestra que los profesores pueden tener distintas expectativas en varios niveles del rango de puntuaciones, y que un profesor 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 para la moderación, el IB no suele seleccionar a alumnos que hayan logrado la máxima puntuación. Esta exclusión busca una mayor equidad, ya que permite que los alumnos con el rango de puntuaciones más elevadas tengan la posibilidad de que sus trabajos se moderen al alza si el profesor ha sido demasiado severo en la corrección. Asimismo, normalmente el IB no selecciona trabajos de alumnos que han obtenido una puntuación de cero.

## Trabajos inusuales o atípicos

Es importante que los profesores corrijan equitativamente el trabajo de sus alumnos. Por tanto, se espera que las puntuaciones y los comentarios del profesor para cada alumno sean representativos del estándar que ha seguido. Esto significa que, casi en todos los casos, se espera que el colegio envíe al IB trabajos si se le solicita para la moderación.

El objetivo de la moderación es comprobar que los profesores corrigen de acuerdo con el estándar global, aun si un trabajo es particularmente difícil de corregir. En esos casos, es posible que el profesor tenga que aportar una justificación más detallada de por qué ha concedido una puntuación concreta. Por ejemplo, si el alumno ha recibido una ayuda adicional, el profesor debe explicarlo en su corrección, y el examinador 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 ajuste de moderación con el trabajo de muestra entregado. Eso sucede si la diferencia entre la puntuación del examinador y la del profesor resulta incoherente, o si examinador considera que el profesor ha sido demasiado generoso o demasiado severo en comparación con el estándar global. En dichos casos, el IB pedirá al colegio que envíe más trabajos para que se pueda aplicar un ajuste de moderación justo. Por ese motivo, el trabajo de todos los alumnos debe estar disponible hasta que se publiquen los resultados.

En ocasiones, esta situación se debe a que varios profesores han corregido los trabajos, pero no han logrado establecer un estándar común. Puesto que se aplica un solo factor de moderación a cada componente de evaluación interna del colegio, es fundamental que los profesores armonicen su estándar de corrección antes de enviar los trabajos al IB.

#### Toma dinámica de muestras

- La moderación con toma dinámica de muestras significa que las puntuaciones de un profesor se aceptan si se encuentran dentro de los márgenes de tolerancia.
- Si la puntuación de un profesor se encuentra fuera de los márgenes de tolerancia, se aplica un factor de moderación aritmético.
- El trabajo de todos los examinadores que comprueban las correcciones de la evaluación interna se supervisa mediante un modelo de calidad similar al que se utiliza en los exámenes.

El principio de la moderación con toma dinámica de muestras es el mismo que se utiliza para los examinadores. Si las puntuaciones que otorga el profesor 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 del margen de tolerancia, entonces se calcula y se aplica un factor de moderación. Esto se explica en el siguiente diagrama.





Figura 51
Diagrama de la moderación con toma dinámica de muestras

A diferencia del enfoque anterior, mediante la moderación con toma dinámica de muestras es mucho menos probable que a los colegios se les aplique un pequeño factor de moderación a las puntuaciones que otorguen, porque, en dichos casos, se considerará que están dentro de los márgenes de tolerancia. Otra diferencia es que el IB podrá proporcionar comentarios más detallados, ya que se requerirá el análisis de un solo examinador como resultado del modelo de calidad que se muestra a continuación.

Con el fin de que la moderación sea justa, los examinadores tienen que comprender el estándar global que establezca el examinador principal y deben realizar la moderación según este estándar. Para garantizar que así sea, el IB utiliza un sistema similar al modelo de calidad que se aplica al trabajo de los examinadores.

Los examinadores 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 "Modelo de calidad", el examinador principal prepara tres tipos de exámenes moderados definitivos:

- Exámenes de práctica: para explicar el estándar global
- Exámenes de aptitud: para comprobar que los examinadores comprenden el estándar global
- Exámenes de control: para comprobar que los examinadores mantienen el estándar global

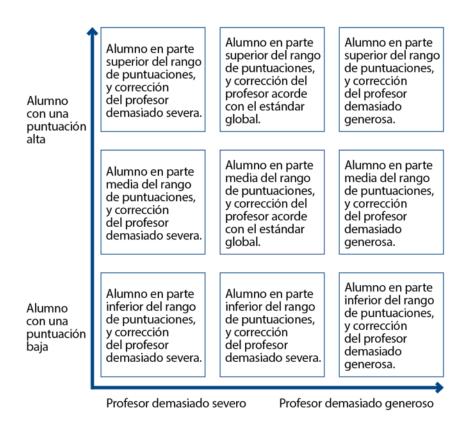
A diferencia del enfoque de moderación anterior, con la toma dinámica de muestras, los examinadores reciben trabajos de distintos colegios sin ningún orden concreto. De esta manera, se evita que los examinadores se formen una opinión sobre si un profesor puntúa demasiado alto o demasiado bajo al ver el trabajo del primer alumno, y que busquen ese patrón en el resto de la muestra del colegio. Además, este método permite incluir exámenes de control sin que resulte obvio que lo son.

El mismo examinador continuará revisando todos los exámenes de un mismo colegio, pero ya no todos juntos. Además, una vez que se haya determinado si es necesario aplicar un factor de moderación, el examinador deberá proporcionar 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. En los cuestionarios de examen, el examinador solo tiene que decidir qué puntuación del rango corresponde a cada respuesta. Sin embargo, en la moderación, el IB también debe asegurarse de que el

examinador sea capaz de determinar en qué casos la corrección del profesor es demasiado severa, demasiado generosa o correcta. En el siguiente diagrama se resume esta situación.

Figura 52
Variedad necesaria de exámenes moderados definitivos



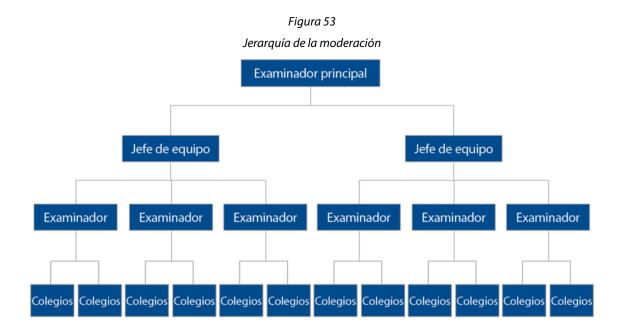
### El sistema anterior: la jerarquía de la moderación

- El sistema de la jerarquía de la moderación se está remplazando por la toma dinámica de muestras, que permite al IB garantizar que todos los examinadores se rijan por el estándar global antes de que procedan a la moderación.
- En el sistema anterior, también se aplicaba un factor de moderación a las puntuaciones de los examinadores (por debajo del examinador principal). El factor de moderación final que se aplicaba a las puntuaciones del profesor era la combinación de todos estos factores. De este modo, cada conjunto de puntuaciones de un profesor estaba en consonancia con el estándar global que estableciera el examinador principal.

Antes de introducir la toma dinámica de muestras, el IB contaba con un segundo enfoque de moderación para comprobar que los examinadores revisaran correctamente las puntuaciones de los profesores. En ese sistema, el examinador principal revisaba una muestra de las puntuaciones que había concedido su equipo de examinadores supervisores y posteriormente estas puntuaciones se ajustaban para garantizar que fueran coherentes con el estándar del examinador principal.

Así, se creaba un proceso jerárquico diseñado para garantizar que las puntuaciones finales otorgadas a cada colegio estuvieran en consonancia con el estándar de corrección que estableciese el examinador

principal. A continuación se ilustran los distintos niveles de jerarquía de un componente de evaluación interna con muchos alumnos:



Este diagrama muestra que:

- Las puntuaciones de un colegio se pueden ajustar en función de la muestra que se entregue a un examinador
- Las puntuaciones de cada examinador se revisan y se ajustan en función de una muestra de su corrección que se entrega a un examinador supervisor (el jefe de equipo)
- A su vez, las puntuaciones del jefe de equipo se pueden ajustar basándose en una muestra de sus correcciones que se entrega al examinador principal

Por lo tanto, se produce una cadena de moderación en la que se puede hacer una serie de ajustes en las puntuaciones de un colegio antes de otorgar una puntuación moderada final (que siga el estándar del examinador principal).

# Comentarios sobre la evaluación interna para los colegios

El propósito de la moderación de la evaluación interna es garantizar que todos los profesores siguen el mismo estándar de corrección. En última instancia, al IB le gustaría que no se tuviera que aplicar ningún factor de moderación a los trabajos de los colegios. Para ayudar a los docentes 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 permite a los profesores 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 los alumnos de la muestra podrían haber conseguido mejores resultados.

## Concesión de calificaciones finales (y totalización)

- En qué consiste la concesión de calificaciones finales
- Las calificaciones finales deben tener el mismo significado, independientemente de la convocatoria en la que el alumno 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, entre ellas la valoración acerca del trabajo de los alumnos por parte de expertos, y los resultados del grupo.
- Lo más importante es la calificación final global que recibe el alumno, 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 "Puntuación y calificación son conceptos distintos". En el proceso de concesión de calificaciones se decide cómo convertir las puntuaciones en calificaciones finales.

Para tomar esta decisión, el examinador principal, el examinador jefe, los examinadores 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 la concesión de calificaciones finales") para llegar a una conclusión que sea justa para los alumnos de ese año, pero también para los alumnos que ya cursaron la asignatura en el pasado. Finalmente, el examinador jefe hace una serie de recomendaciones al IB sobre cuáles deberían ser los resultados para esa convocatoria de exámenes.

Durante el proceso de concesión de calificaciones finales también se reúnen pruebas sobre la eficacia de las evaluaciones. Dicha información se utiliza para mejorar las convocatorias de exámenes posteriores. Concretamente, en el primer año de evaluación de un nuevo curso, el responsable del currículo del IB — que está encargado del desarrollo del currículo del curso— participa en la reunión de concesión de calificaciones con un doble fin: por una parte, facilitar la coherencia entre los resultados de la evaluación y la intención de los objetivos generales del curso, y, por otra parte, dar forma al futuro desarrollo del currículo.

El objetivo más destacado de estas discusiones es establecer los límites de calificación, donde se define la cantidad mínima de puntos que un alumno 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 que recomienden varios *límites de calificación estimativos* y calcula el resto de los límites aritméticamente. En el PAI, el PD y el POP, en los que hay siete calificaciones, los límites de calificación estimativos son 2-3, 3-4 y 6-7. Estos límites representan el máximo nivel de conocimiento y comprensión de una asignatura (6-7), el desempeño de un alumno con conocimiento y comprensión básicos (2-3), y un nivel sólido de conocimiento y comprensión (3-4).

Los límites de calificación deben cambiar en cada convocatoria de exámenes. Esto se debe a que las preguntas que se plantean a los alumnos cambian, 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 cuestionarios de examen tengan la misma dificultad cada año. Sin embargo, dada la importancia de lo que hay en juego, no se realizan pruebas previas de los exámenes, debido al riesgo de que se divulguen las preguntas.

En el proceso de concesión de calificaciones, los examinadores y el IB tienen en cuenta los siguientes aspectos:



Figura 54
Aspectos que se deben tener en cuenta durante el proceso de concesión de calificaciones finales

Grupo	• ¿Cómo es el grupo actual que realiza el examen en comparación con grupos de años anteriores?
Examen	• ¿En qué medida el examen ha cumplido las expectativas que tenía el autor? • ¿Qué opinaron los profesores sobre la dificultad del examen?
Exámenes escritos de los alumnos	• El trabajo de los alumnos se analiza y se compara con el de años anteriores y con los descriptores de calificaciones finales.
Resultados	• ¿Cómo son los resultados de los alumnos en comparación con los de años anteriores?
Equilibrio	• ¿Todas las pruebas disponibles llevan a una sola conclusión, o son contradictorias?
Casos especiales	Se analizan los casos de aquellos alumnos que han dado respuestas poco comunes o que tienen circunstancias especiales.

En muchas ocasiones, el IB cuenta con varios componentes de evaluación que se combinan para alcanzar una calificación general. En la sección "Totalización", que aparece más adelante, se ofrece más información sobre este tema. El aspecto fundamental que los examinadores deben recordar durante la concesión de calificaciones finales es que lo más importante es el resultado total. Si es necesario, 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 de los alumnos, en qué puntos hay un cambio en la calidad del trabajo que justifique pasar a otro descriptor de calificaciones finales
- Para cada componente, determinar las puntuaciones que conformarán los límites de calificación
- Asegurarse de que el conjunto de los límites de calificación de cada curso da como resultado una concesión de calificaciones justa
- Decidir qué calificación conceder a las respuestas de los alumnos que sean atípicas o inusuales

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

- La valoración de las respuestas de los alumnos y las pruebas obtenidas de los resultados coinciden ampliamente, teniendo en cuenta la información sobre el grupo
- Se pueden explicar las variaciones que haya en el desempeño de los colegios
- El examinador jefe y el director 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 jefe a partir de las discusiones mantenidas durante el proceso de concesión de calificaciones. En el PD, el POP y el PAI, los límites de calificación estimativos son 2-3, 3-4 y 6-7. Estos límites representan el máximo nivel de conocimiento y comprensión de una asignatura (6-7), el desempeño de un alumno con conocimiento y comprensión básicos (2-3), y un nivel sólido de conocimiento y comprensión (3-4).

El resto de límites de calificación (1-2, 4-5 y 5-6) se calculan aritméticamente basándose en los límites estimativos, por lo que se denominan límites de calificación interpolados.

Estos límites de calificación interpolados se basan en pruebas que se obtienen al revisar el trabajo de los alumnos, y tienen una función importante al considerar los logros generales de los alumnos. Si hay un cambio significativo en la proporción de alumnos que reciben una calificación de 5 o de 1, hará falta discutir los resultados teniendo en cuenta el contexto del grupo de alumnos, y esto puede conducir a que se

reconsideren los límites estimativos o, en circunstancias excepcionales, a revisar los trabajos de los alumnos que se encuentren en esos límites.

## El impacto de la evaluación electrónica en la concesión de calificaciones finales

Sea cual sea la forma de evaluación utilizada, el proceso de concesión de calificaciones se debe llevar a cabo. La introducción de la evaluación electrónica no cambiará en absoluto los principios de la concesión de calificaciones.

Sin embargo, es posible que la evaluación electrónica afecte al proceso de concesión de calificaciones en el sentido de que permite que las evaluaciones midan más eficazmente si se han alcanzado los objetivos del curso. Muchas veces, los examinadores tienen dificultades para encontrar pruebas de análisis, investigación o pensamiento colaborativo en los exámenes en papel, lo que significa que estos aspectos de los descriptores de calificaciones finales están infrarrepresentados. La evaluación electrónica puede ayudar a resolver este problema.

## Pruebas utilizadas en la 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 de los profesores sobre la evaluación
- El juicio experto de los examinadores sobre los ejemplos de trabajos de los alumnos
- El 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 alumnos del año en curso

La primera tarea del proceso de concesión de calificaciones finales es plantearse en qué medida el grupo de alumnos 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 cuestionarios de examen. Por el contrario, si hay muchos colegios que inscriben a alumnos en una asignatura por primera vez, o si hay una proporción elevada de alumnos 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 los examinadores tendrán en cuenta al comparar distintos grupos de alumnos que realizan la evaluación, se pueden indicar:

- Cambios en el número de alumnos que realizan las evaluaciones
- Cambios en la proporción de alumnos que realizan las evaluaciones en español, francés e inglés (y en otras lenguas según corresponda)
- El número de nuevos colegios y la cantidad de alumnos que haya en esos colegios
- Los cambios en las opciones que elijan los alumnos

#### 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 alumnos merecedores de una calificación de 6 y



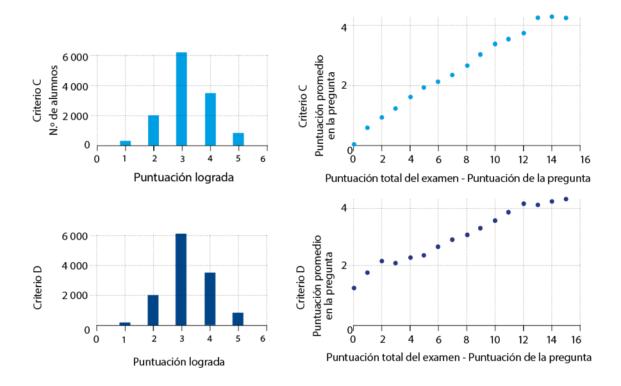
los merecedores de una calificación de 7 no se logra este propósito, entonces los examinadores supervisores tendrán esto en cuenta cuando determinen los límites de calificación.

Los examinadores supervisores se basarán en los comentarios de los profesores sobre los cuestionarios de examen (principalmente los formularios G2) y en las estadísticas sobre el desempeño en preguntas o elementos concretos (véase el diagrama que se muestra a continuación). También utilizarán resúmenes de informes de distintos examinadores.

Figura 55 Ejemplo de las estadísticas que los examinadores tienen a su disposición acerca de las preguntas

Resumen de las estadísticas sobre las preguntas						
Pregunta	Preguntas que se han intentado contestar	Total de alumnos para el examen	Proporción de alumnos que han intentado contestar	Puntuación promedio en la pregunta	Puntuación más alta conseguida en la pregunta	Desv. estándar en la puntuación de la pregunta
01 Criterio A	12.071	42.042	28,71 %	2,7	5	0,95
01 Criterio B	12.071	42.042	28,71 %	2,6	5	0,94
01 Criterio C	12.071	42.042	28,71 %	2,8	5	0,84
01 Criterio D	12.071	42.042	28,71 %	3,2	5	0,79
02 Criterio A	29.848	42.042	71,00 %	2,8	5	0,98
02 Criterio B	29.848	42.042	71,00 %	2,7	5	0,93
02 Criterio C	29.848	42.042	71,00 %	2,9	5	0,83
02 Criterio D	29.848	42.042	71,00 %	3,2	5	0,79

Correlación entre la pregunta y el resto del componente				
Pregunta	Coeficiente de correlación			
01 Criterio A	0.8044			
01 Criterio B	0.8018			
01 Criterio C	0.7983			
01 Criterio D	0.7253			
02 Criterio A	0.7954			
02 Criterio B	0.7944			
02 Criterio C	0.7865			
02 Criterio D	0.7320			



Los examinadores supervisores también se apoyarán en su propia experiencia de corregir trabajos de alumnos durante la convocatoria de exámenes. Por ese motivo, es importante que en las discusiones sobre la concesión de calificaciones participen personas que hayan corregido cada uno de los componentes durante la convocatoria. Aunque preside estas discusiones, no es obligatorio que el examinador jefe haya corregido trabajos de alumnos; de hecho, puede que sea útil no tener ideas preconcebidas. Sin embargo, esto impone a los examinadores principales la responsabilidad añadida de tener que proporcionar la información necesaria sobre la calidad general de los trabajos de los alumnos.

#### Búsqueda de pruebas en los exámenes

Los descriptores de calificaciones finales establecen las características que se esperan ver en los trabajos de los alumnos para cada calificación. Durante esta fase del proceso de concesión de calificaciones, los examinadores no deben centrarse en las puntuaciones otorgadas, sino en la naturaleza de las respuestas de los alumnos y en qué medida encajan con los descriptores de calificaciones finales. Por ese motivo, en la medida de lo posible, se seleccionan exámenes en los que la calidad de las respuestas sea semejante a lo largo de toda la prueba, en vez de exámenes en los que se hayan obtenido puntuaciones altas en algunas preguntas y bajas en otras.

Antes de empezar a revisar los exámenes con las respuestas de los alumnos, a menudo es útil mirar algunos ejemplos de años anteriores para recordar a los examinadores las expectativas correspondientes a las distintas calificaciones. Los exámenes que se encuentran en el límite de calificación deben estar disponibles durante todo el proceso de concesión de calificaciones finales.

Antes de la reunión de concesión de calificaciones, el equipo de examinadores supervisores, y en concreto los examinadores 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 cuestionario de examen, y la distribución general de puntuaciones, el responsable de la asignatura del IB sugiere el rango de



puntuaciones (exámenes de muestra) que los examinadores supervisores necesitan para empezar a revisar los trabajos de los alumnos para cada calificación.

Cada examinador supervisor debe revisar los exámenes escritos de los alumnos que se hayan seleccionado y determinar qué descriptor de calificaciones finales refleja mejor la calidad del trabajo. Es importante centrarse en el examen en conjunto y no dejarse influir por las puntuaciones otorgadas. En este sentido, es útil saber qué preguntas se diseñaron para proporcionar pruebas de las cualidades correspondientes a la calificación más alta, pero dichas preguntas simplemente deben ayudar a tomar una decisión general. Esta tarea resulta complicada, ya que incluso los alumnos que mantengan en su examen un nivel relativamente homogéneo mostrarán a menudo en sus respuestas cualidades correspondientes a calificaciones muy variadas. Por tanto, los examinadores tendrán que juzgar cuál es la calificación más adecuada. Es aceptable, y a menudo útil, indicar si un examen corresponde plenamente al descriptor de una calificación final o si está cerca de la siguiente calificación (esto se suele indicar de la siguiente manera: "7-" o "6+", por ejemplo).

Para llevar a cabo su cometido, es esencial que los examinadores 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 (excepto al examinador jefe) acerca de los resultados del análisis estadístico hasta que hayan finalizado esta etapa.

Determinar qué calificación final corresponde a un examen es un ejercicio muy subjetivo, y es probable que haya variaciones entre los distintos examinadores. Las puntuaciones y las calificaciones representan desempeños sutilmente distintos. Por este motivo, sería razonable que un alumno que haya demostrado una buena comprensión aunque haya cometido varios errores reciba una calificación más alta que un alumno 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 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 alumnos, y todas las decisiones deben tener esto en cuenta.

Hay investigaciones que indican que los examinadores 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 "Otras formas de corrección"). Por esto, el IB pide a los examinadores 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 supervisores para conceder las calificaciones finales, los resultados 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 para la que hay pruebas razonables para colocar el límite de calificación hasta la puntuación más alta. No existe una definición precisa de cómo establecer una zona de incertidumbre, así que los examinadores supervisores deben llegar a un acuerdo al respecto en función de sus decisiones individuales.

Figura 56 Ejemplo de resultados de la revisión de exámenes de alumnos. En un proceso de concesión de calificaciones real, se utilizarían más exámenes de cada puntuación.

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+	<b>1</b> %
4	50	4-	4	4	3+	E
5	49	4-	4-	4-	4-	Zona de incertidumbre
6	49	3+	3	4-	3+	e inc
7	48	4-	3+	4-	3+	na d
8	48	3+	4-	4-	3+	1 <sub>8</sub>
9	47	3	3	3+	3	
10	47	3	3	3	4-	

La única excepción a este proceso se da con los cuestionarios de examen de opción múltiple. La experiencia demuestra que emitir juicios sobre los límites de calificación en función de la calidad del trabajo del alumno es muy difícil en aquellas pruebas de examen que solamente incluyen preguntas de opción múltiple. Esto se debe a que las respuestas no muestran qué han hecho realmente los alumnos y, por tanto, se carece de base suficiente para tomar una decisión. En ese tipo de pruebas, los límites de calificación se calculan de modo que el porcentaje de alumnos de cada calificación final sea el más cercano posible al obtenido a partir de una estimación en la prueba con la 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, los examinadores supervisores deberían poder establecer los límites de calificación simplemente teniendo en cuenta las preguntas del examen y las respuestas que se esperan de los alumnos. Sin embargo, en la realidad es muy difícil que dichas opiniones puedan formularse con precisión antes de conocer cómo han respondido realmente los alumnos. Cresswell (2000) llegó a la conclusión de que los correctores solían identificar correctamente si los cuestionarios de examen resultaban más fáciles o más difíciles de una convocatoria a otra, pero no lograban estimar bien en qué medida eran más fáciles o más difíciles.

[...] hay buenas razones teóricas y empíricas para creer que "mantener los estándares según la definición de enfoque flexible basado en criterios" es demasiado complejo incluso para los correctores experimentados. Además, hay pruebas empíricas de que los juicios de evaluación de los correctores se ven influidos por factores que no les deberían afectar, como la coherencia de puntuaciones en un mismo examen de un alumno. Estos son buenos motivos para dudar de que las calificaciones finales que deciden estos profesionales basten, por sí solas, para mantener los juicios que se emiten sobre los exámenes.

Baird, Cresswell y Newton, 2000.

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 alumnos con las mismas capacidades y los mismos logros alcanzados, que han acudido a colegios parecidos con políticas de admisión idénticas, que han tenido profesores 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.

Para facilitar su labor de concesión de calificaciones finales, se proporciona a los examinadores supervisores los límites recomendados estadísticamente. Estos son límites de calificación que ofrecerían exactamente el mismo porcentaje acumulado que el año anterior hasta esa calificación. El término porcentaje acumulado se refiere al número total de alumnos que recibió una calificación concreta o alguna superior. Este porcentaje se utiliza para explicar el hecho de que, si los mejores alumnos obtienen mejores resultados (de manera que aumenta la proporción de alumnos que reciben una calificación de 7), se puede esperar que haya menos alumnos con una calificación de 6.

Además, a los examinadores que participan en el proceso de concesión de calificaciones finales se les proporciona información sobre la puntuación media (el promedio) que obtienen los alumnos, 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.

Volviendo a las dos citas anteriores, es importante recordar que el grupo de alumnos que realizan los exámenes cambia en cada convocatoria. Por ese motivo, es probable que los postulados de la segunda cita no se puedan cumplir del todo. Así sucederá especialmente si se trata de dos grupos muy distintos entre sí. De la misma manera, cuando hay una gran cantidad de alumnos con experiencias educativas similares (por ejemplo, que provengan de colegios autorizados por el IB), es más probable que las diferencias considerables en los resultados se deban a que los alumnos 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 alumnos. 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 alumnos, se puede esperar que haya una variación considerable en el resultado global, pero es relativamente poco probable que esto suceda si hay miles de alumnos.

#### Cómo encontrar el equilibrio entre las pruebas

Ningún tipo de prueba es más importante que otro. La tarea del examinador jefe y su equipo es tenerlas todas en cuenta para poder hacer una recomendación.

Similitud del grupo Pruebas que **Perspectivas** aportan **Estadísticas** sobre los sobre el exámenes los desempeño resultados en la evaluación Límites de calificación recomendados

Figura 57
Pruebas que respaldan la selección de los límites de calificación

Con frecuencia, las distintas pruebas apuntan a los mismos resultados (por ejemplo, que 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. Teniendo en cuenta que lo importante es el resultado total y no los componentes individuales, el examinador jefe puede comparar la calificación media del grupo con las predicciones de los profesores, 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 examinador jefe tiene que enviar sus recomendaciones, acompañadas de las justificaciones necesarias, al director en jefe de evaluación del IB.

#### Descriptores de calificaciones finales

Los descriptores de calificaciones finales son un resumen de las características de desempeño para cada calificación. Su papel es fundamental para establecer los límites de calificación, ya que determinan qué deben buscar los examinadores al revisar el trabajo de los alumnos. Esto significa que se deben redactar descriptores para cada asignatura, con el fin de que se apliquen de manera coherente de una convocatoria de exámenes a otra.

La importancia de los descriptores de calificaciones finales es tal que se consideran una parte esencial de la validez de la evaluación de los cursos del IB. Estos descriptores tienen que reflejar los objetivos generales y específicos de la asignatura para garantizar que el IB evalúa y recompensa lo que se pretende.

Además, los descriptores actúan como referencia común para todas las asignaturas, con lo que fomentan la equiparabilidad entre estas. Así pues, es necesario contar con descriptores generales de calificaciones finales para que, por ejemplo, una calificación de 4 signifique lo mismo en todas las asignaturas.



Estos principios implican que debe haber una relación clara entre los objetivos del programa del IB y los descriptores generales de calificaciones finales, así como una evolución lógica hacia los descriptores de calificaciones finales específicos de cada asignatura. Esta idea se ilustra en el siguiente diagrama.

Figura 58 Relación entre los descriptores generales de calificaciones finales y los descriptores de calificaciones finales de cada asignatura



## Límites de calificación fijos

- Cuando las mismas tareas se repiten en cada convocatoria de exámenes, se espera que los límites de calificación tampoco sufran cambios.
- Si hay pruebas de que los estándares son incorrectos o han cambiado con el tiempo, el IB revisará y cambiará estos límites "fijos".

En la mayor parte de la evaluación interna, y también en algunos componentes de evaluación externa, se pide a los alumnos que realicen 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 principal y al examinador 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 modificar 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 destacable en el comportamiento de los alumnos (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 de los alumnos, o una revisión del enfoque del IB respecto a la corrección o a la moderación que modificase el estándar de corrección.

#### Totalización

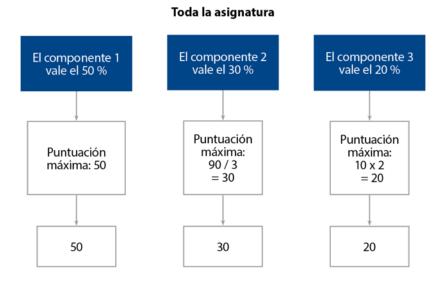
- La totalización es el proceso de combinar las puntuaciones obtenidas en los componentes para obtener un resultado total.
- Para que cada componente contribuya a la puntuación final en la medida que se desea (lo que se conoce como *porcentaje del total de la evaluación* o, en ocasiones, *ponderación*), puede que sea necesario ajustar proporcionalmente las puntuaciones de los componentes.

- El IB utiliza un *modelo de compensación* mediante el cual los alumnos pueden compensar un mal desempeño en un componente o pregunta si obtienen un buen resultado en otro.
- La calificación final que un alumno 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.

Figura 59
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 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.

Hay que recordar que no es necesario que las puntuaciones individuales del alumno coincidan con el porcentaje del total de la evaluación del componente. El IB es consciente de que no todos los alumnos tienen los mismos puntos fuertes, por lo cual se utilizan diversos instrumentos de evaluación. Es importante comparar los porcentajes del total de la evaluación reales de los resultados del grupo entero de alumnos con los porcentajes que se establecieron al diseñar la asignatura. Si, en una convocatoria de exámenes, la contribución de un componente es mucho mayor de lo que se esperaba, eso puede ser señal de que el examen era especialmente fácil y que se deben utilizar otras pruebas del desempeño de los alumnos para poder conceder las calificaciones finales. Por otro lado, si la contribución de un componente es



habitualmente mucho mayor de lo que se esperaba, entonces se deberá revisar el diseño, ya que eso indicaría que la evaluación no funciona tal como se había previsto.

Es posible que este enfoque no refleje los métodos más sofisticados de ponderación, de combinación (totalización) y de ajuste proporcional que describió, por ejemplo, Wood (1991, capítulo 10), pero emplea los principios sólidos de la evaluación basada en criterios y respalda los principios de evaluación del IB.

El uso de la totalización para obtener la puntuación final implica que un alumno puede compensar un mal resultado en un componente con un buen resultado en otro, ya que lo único que afecta a la calificación final es la puntuación total. Este concepto se denomina *modelo de compensación*, en contraposición al *modelo de dominio*, en el que el alumno tiene que demostrar el nivel requerido en todos los componentes para poder obtener una calificación final concreta.

Merece la pena destacar que la calificación final que un alumno obtiene en una asignatura se determina a partir de la totalización de las puntuaciones logradas en cada componente, y no de calificaciones por componente. Como cada calificación en un componente representa un rango de puntuaciones, es posible que dos alumnos con las mismas calificaciones en los componentes reciban distintas calificaciones en la asignatura.

# Los controles de calidad en la concesión de calificaciones finales y los informes de distribución

- Hay distintos procesos que contribuyen a establecer los límites de calificación y cuya intención es ayudar al IB a tomar una decisión equilibrada.
- El examinador jefe y su equipo de examinadores supervisores hacen recomendaciones sobre los límites de calificación.
- El responsable de la asignatura del IB se encarga de comprobar que en la concesión de calificaciones finales se hayan seguido los principios y procesos del IB.
- Las recomendaciones y las justificaciones (incluidos los datos clave) se registran en un informe de la asignatura para distribución en 24 horas, que examina minuciosamente el equipo directivo de la evaluación del IB.

El objetivo del proceso de concesión de calificaciones finales es respaldar las valoraciones que emiten los examinadores, para lo cual se reduce la posibilidad de que haya sesgos involuntarios y de que los examinadores tomen decisiones basándose en trabajos atípicos. El responsable de la asignatura del IB debe supervisar este proceso con el fin de que se siga correctamente y de ayudar a los examinadores supervisores a identificar las dificultades que aparecerán.

El examinador jefe, junto a su equipo de examinadores supervisores, es el encargado de realizar las recomendaciones iniciales sobre dónde se deben establecer los límites de calificación. Estas recomendaciones deben contar con el respaldo de un argumento sólido en el que se explique por qué estos límites son los más adecuados según las pruebas disponibles (que a veces serán contradictorias).

El responsable de la asignatura del IB aportará una perspectiva crítica y constructiva durante este proceso, con el fin de garantizar que las recomendaciones del examinador jefe sean equilibradas y estén justificadas.

A continuación, estas recomendaciones y justificaciones se presentan en el informe de la asignatura para su distribución en 24 horas. Este informe también contiene datos subyacentes, como los cambios en el grupo de alumnos 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 (normalmente el jefe del programa y el jefe de principios y prácticas de evaluación) para determinar si los argumentos son lo suficientemente sólidos.

En caso de que el equipo directivo de la evaluación del IB no esté conforme con las recomendaciones, discutirá sus inquietudes con el examinador jefe 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 en jefe de evaluación del IB tiene la última palabra sobre dónde se establecen los límites de calificación, basándose en las recomendaciones que recibe de los examinadores jefe.

# Otorgamiento del título académico de un programa

Los títulos académicos de los programas del IB (el diploma del IB, el certificado del POP y el certificado del PAI) 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".

#### Profesores observadores

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 a los profesores a asistir como observadores a las reuniones de concesión de calificaciones (los detalles se concretarán en función de si son reuniones presenciales o virtuales). Se espera que los profesores observadores expliquen a sus compañeros lo que han presenciado y ofrezcan un informe para la comunidad de la Red de educadores del IB (IBEN). Para obtener más información, escriba a support@ibo.org.

## Principios de la concesión de calificaciones finales

Los principios subyacentes de la concesión de calificaciones finales del IB son los siguientes:

- 1. 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 de los alumnos para establecer estos límites de calificación, se utilizarán las pruebas disponibles para establecer el límite que resulte más adecuado.
- 2. A continuación, los demás límites de calificación se determinan aritméticamente siguiendo el procedimiento adecuado.
- 3. 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, de las pruebas estadísticas y de información sobre el grupo de alumnos. Todas estas pruebas se deben tener en cuenta por igual y se debe alcanzar un punto de equilibrio.
- 4. Si el grupo de alumnos 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:
  - a. Los grupos de alumnos suelen variar de un año a otro, sobre todo en las asignaturas con pocos alumnos.
  - Si los resultados varían respecto a los de años anteriores, se espera que haya pruebas claras que expliquen las razones.
- 5. 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.
  - a. Todos los límites de calificación pueden cambiar de un año a otro, incluso para las tareas de evaluación interna.
  - b. El IB procura que el nivel de dificultad de las evaluaciones sea coherente de un año a otro, pero dicho nivel puede variar en determinadas pruebas.
- 6. La prioridad son los límites de calificación globales por asignatura, porque son los que afectan al alumno 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 de casos en riesgo

- Lo más importante que se obtiene de la evaluación es la calificación que recibe el alumno; por ese motivo, los controles finales se centran en ella.
- El propósito de estos controles finales es buscar cualquier elemento inusual. Toda inquietud con respecto a la corrección se habrá abordado antes de llegar a esta etapa.
- Si durante este proceso se identifica algún patrón de cambios a las puntuaciones, el IB investigará el caso y posiblemente se revise la corrección de todo el trabajo realizado por ese examinador concreto.
- Los encargados de realizar estos controles son los examinadores más coherentes y, en caso de que haya alguna diferencia, su puntuación se considera más adecuada que la anterior.

El modelo de calidad para la corrección consiste en comprobar periódicamente el estándar de corrección de los examinadores mediante exámenes de control, con el fin de que el IB compruebe que los exámenes se corrigen correctamente. Otra manera de revisar la calidad es buscar resultados inesperados y comprobar si son correctos. Esto se denomina *revisión de casos en riesgo*.

Actualmente se emplean dos criterios para identificar resultados inusuales:

- Casos en los que un alumno ha recibido una calificación mucho más baja de lo que había previsto el profesor.
- Casos en los que los resultados totales de un colegio difieren mucho de los del año anterior. El IB se centra especialmente en aquellos casos en los que un componente es muy distinto a como había sido en años anteriores.

Esto no significa, en ninguno de los dos casos, que las puntuaciones otorgadas sean incorrectas, pero sí indica que hay áreas en las que sería adecuado realizar un control de calidad adicional. En general, se da prioridad a los casos en los que hay pruebas de que los alumnos han obtenido peores resultados de lo esperado, más que a los que han obtenido resultados mejores de lo previsto.

Estos controles de calidad se confían a aquellos examinadores que, a través de los exámenes de control, han demostrado corregir de la manera más coherente, lo cual permite confiar en que las puntuaciones que otorgan son adecuadas. Por este motivo, si estos examinadores recomiendan una puntuación distinta, se aplicará su decisión en lugar de la del examinador que corrigiera el trabajo en primer lugar. Sin embargo, si se detecta una diferencia considerable entre dos examinadores coherentes, se investigará más a fondo y es posible que se busque una tercera opinión para comprender los motivos de dicha diferencia.

En la corrección electrónica, la computadora registra todas las puntuaciones, pero en los exámenes en papel es posible que los colegios vean dos o más conjuntos de puntuaciones debido a este proceso de revisión. El IB siempre explicará qué puntuaciones son las del examinador con mayor fiabilidad y nivel en la jerarquía.

La corrección de casos en riesgo tiene un doble propósito:

- 1. Comprobar que los alumnos que tuviesen un resultado inusual reciban un resultado justo.
- Buscar pruebas que revelen problemas sistemáticos con la corrección (por ejemplo, una pregunta que varios examinadores hayan tenido dificultades para corregir, o una falta de coherencia de varios examinadores al corregir). En estos casos, se revisaría la corrección del trabajo de los alumnos afectados.



#### El Comité de la evaluación final

- El Comité de la evaluación final es el órgano que toma las decisiones estratégicas en cada convocatoria de exámenes.
- Está formado por personal sénior de varias divisiones del IB y examinadores jefe de distintas áreas disciplinarias.
- Además de dar su aprobación final para la publicación de resultados, también estudia casos de conducta improcedente, las solicitudes de consideración especial y cualquier otra cuestión pertinente.
- Además de los miembros con derecho a voto, se puede invitar a colegios y otras partes interesadas a observar sus reuniones.

El Comité de la evaluación final representa la última etapa en la aprobación de los resultados de los alumnos. Se trata del órgano que revisa las recomendaciones del director en jefe 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. Además, también establece 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 de las divisiones de evaluación, académica y de servicios a los colegios
- Examinadores jefe de distintas áreas disciplinarias

A diferencia del personal sénior de la división de evaluación, que realiza controles en los informes de la asignatura para distribución en 24 horas, el Comité de la evaluación final actúa como órgano de supervisión. Este comité revisa los resultados a nivel global (por ejemplo, el porcentaje de alumnos que completa cada programa) y cualquier cuestión que le presente el personal de la división de evaluación del IB.

De manera similar, el Comité de la evaluación final revisa las cuestiones de conducta improcedente y mala administración. Un subcomité realiza las debidas recomendaciones, que se discuten y sirven para establecer precedentes en los casos nuevos.

Las solicitudes de consideración especial se tratan de modo similar: el personal de la división de evaluación del IB hace recomendaciones al Comité de la evaluación final, y estas se discuten antes de llegar a una decisión final.

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.

#### Conflictos de interés

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 media de precaución se repite al principio de cada reunión.

#### Observadores

El IB fomenta que haya observadores en las reuniones del Comité de la evaluación final, con el fin de:

- Hacer que los procedimientos del Comité de la evaluación final sean más transparentes
- Permitir que se aporten comentarios sobre los procedimientos del Comité de la evaluación final
- Reconocer la colaboración entre las distintas partes que constituyen el IB
- Ofrecer la posibilidad de aportar sugerencias de cambios y de mejoras en los procedimientos del Comité de la evaluación final, por medio de un informe escrito

Para que el IB pueda cumplir estos objetivos, se pide a los observadores que envíen un informe sobre su participación al director en jefe de evaluación 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 de la evaluación final, los observadores también están sujetos a las debidas restricciones relativas a la confidencialidad y a los conflictos de interés. Los observadores no tienen derecho a voto en las decisiones del comité.

Para obtener más información sobre cómo ser observador en una reunión del Comité de la evaluación final, escriba a support@ibo.org.



## Preparación para la publicación de los resultados

- Una vez corregidos todos los exámenes y establecidos todos los límites de calificación, se deben completar varios procesos para que los resultados de los alumnos estén listos para publicarse.
- En los casos en los que no haya pruebas del trabajo de un alumno, el IB puede calcular una puntuación mediante el procedimiento de ausencia de puntuación.
- Los resultados de los alumnos se deben combinar para determinar si pueden recibir el diploma del IB, el certificado del POP o el certificado del PAI.
- Se revisarán todos los casos de alumnos a los que se les apliquen consideraciones especiales.
- La publicación de resultados exige establecer un sistema eficaz de control de cambios, con el fin de detectar cualquier alteración en las calificaciones e informar al respecto a las partes interesadas pertinentes, que principalmente serán colegios y universidades.

Entre la etapa de establecimiento de los límites de calificación y la de publicación de los resultados para los alumnos hay varios procesos y procedimientos que se deben completar. Algunos solo son necesarios para determinados alumnos, como los que cursan asignaturas en la categoría Anticipado o para los que se han tenido que aplicar consideraciones especiales, y otros de los procesos y procedimientos afectan a todos los alumnos.

## Casos de ausencia de puntuación

- Cuando el IB no pueda acceder al trabajo de un alumno debido a motivos ajenos a la voluntad de este o del colegio, se estimará una puntuación con el fin de minimizar los posibles efectos negativos para el alumno.
- Dicha estimación se debe basar en pruebas, y si los trabajos del alumno no son suficientes para realizar la estimación, entonces se debe encontrar otra solución para sustituir la puntuación ausente.
- El proceso de ausencia de puntuación se basa en los resultados medios de los alumnos, por lo que esta estimación beneficiará aproximadamente al mismo número de alumnos a los que perjudicará. El resultado más justo siempre será aquel que se obtenga al corregir el trabajo real del alumno.

En ocasiones, el IB no dispone del trabajo de un alumno y no lo puede corregir, por motivos que escapan del control del colegio y del alumno. Esto sucede, por ejemplo, cuando el servicio postal pierde los exámenes con las respuestas de los alumnos, o cuando un alumno 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 que el alumno no se vea injustamente perjudicado.

Este procedimiento está pensado para aquellas circunstancias en las que la ausencia del trabajo del alumno se debe a alguna acción del IB o de terceros (a excepción del colegio), a consecuencia de la cual no sería razonable exigir al alumno que completara la evaluación en otra ocasión.

El procedimiento de ausencia de puntuación no pretende sustituir a las adecuaciones de acceso ni a las consideraciones especiales (salvo en casos excepcionales).

Este procedimiento siempre se debe basar en pruebas que demuestren los logros del alumno. Por este motivo, si no se 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 han alcanzado todos los alumnos en el componente afectado, que se comparan con los resultados obtenidos en los demás componentes. Hay que reconocer que, al utilizar dicha media, se espera que la cantidad de alumnos cuya puntuación empeore debido a esta estimación sea aproximadamente la misma que de alumnos cuya puntuación mejore. Por lo tanto, el resultado siempre será más justo si se puede corregir el trabajo real del alumno.

Funcionamiento del procedimiento de ausencia de puntuación Puntuación Puntuación Usar Calcular cálculo Puntuación Puntuación **Puntuación** Puntuación Componente Componente ausente ausente Puntuación Puntuación Componentes Componentes corregidos corregidos (

Figura 60

Media de todos los alumnos

Cálculo para el alumno afectado

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. Así sucede en el PAI, por lo que en este programa se utiliza un enfoque específico para determinar la calificación ausente.

## Resultados en los programas

- El IB otorga los títulos de los programas siempre que se cumplan ciertos criterios precisos, y no basándose en valoraciones individuales.
- Cada programa se rige por unos criterios propios para la obtención del certificado correspondiente.

Después de conocer los resultados de un alumno en todas las asignaturas, se realiza un cálculo para determinar si puede recibir el diploma del IB, el certificado del POP o el certificado del PAI. Para lograrlo, el alumno debe cumplir los criterios que se explican en sus correspondientes secciones.

#### Modo de calcular el resultado del PD

Para calcular la puntuación total lograda en el PD, 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 de los componentes troncales. Así, la puntuación más alta posible que puede obtener un alumno es 45 puntos\*.

Este enfoque implica que las asignaturas de NM y NS tienen el mismo peso a la hora de determinar la puntuación final del alumno.

\* La puntuación máxima de 45 se obtiene de la siguiente manera: 6 (asignaturas) multiplicado por 7 (calificación máxima), más 3 (máximo de puntos de los componentes troncales).



#### Matriz de puntos para los componentes troncales

A diferencia de las demás asignaturas, Teoría del Conocimiento (TdC) y la Monografía reciben calificaciones alfabéticas, que van de la A a la E. El tercer componente troncal (Creatividad, Actividad y Servicio) no recibe una calificación, ya que no sería pertinente evaluar el desempeño en esta área.

Los componentes troncales aportan entre 0 y 3 puntos a la puntuación total lograda en el PD. Si el alumno obtiene una calificación de E en TdC o en la Monografía, o si no completa Creatividad, Actividad y Servicio, es posible que no obtenga el diploma. Para calcular la puntuación, se utiliza la tabla que aparece a continuación.

Figura 65 Matriz de puntos para los componentes troncales

	Teoría del Conocimiento (TdC)					
	Calificación final obtenida	А	В	С	D	Е
,e	Α	3	3	2	2	
Monografía	В	3	2	2	1	Condición excluyente
Mon	С	2	2	1	0	
	D	2	1	0	0	
	E	Condición excluyente				

#### **Condiciones excluyentes**

Un alumno no podrá recibir el diploma si se da alguna de las nueve condiciones siguientes:

- No se han cumplido los requisitos de CAS.
- La puntuación total no alcanza 24 puntos.
- Se ha obtenido una "N" (no se ha concedido ninguna calificación) en Teoría del Conocimiento, la Monografía o alguna de las asignaturas que cuentan para la obtención del diploma.
- Se ha obtenido la calificación "E" (elemental) en la Monografía, Teoría del Conocimiento o ambas.
- Se ha obtenido una calificación de 1 en cualquier asignatura y nivel.
- Se ha obtenido una calificación de 2 en tres ocasiones o más (NM o NS).
- Se ha obtenido una calificación de 3 o inferior en cuatro ocasiones o más (NM o NS).
- Se han obtenido menos de 12 puntos en las asignaturas de NS (para los alumnos matriculados en cuatro asignaturas de NS, contarán las tres calificaciones más altas).
- Se han obtenido menos de 9 puntos en las asignaturas de NM (los alumnos matriculados en dos asignaturas de NM deberán obtener al menos 5 puntos en este nivel).

#### Diploma bilingüe

En lugar del diploma del IB habitual, el alumno puede obtener un diploma bilingüe si cumple al menos uno de los siguientes criterios:

- Haber completado dos lenguas del Grupo 1 con un mínimo de una calificación final de 3 en ambas.
- Haber completado una asignatura del Grupo 3 o del Grupo 4 en una lengua distinta de su lengua del Grupo 1. El alumno debe obtener una calificación final de 3 o superior tanto en la lengua del Grupo 1 como en la asignatura del Grupo 3 o del Grupo 4.

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
- Un programa de estudios del colegio
- Una asignatura adicional a las seis asignaturas conducentes al diploma

#### Modo de calcular el resultado del POP

No existe una puntuación para el certificado del POP.

Si el alumno cumple todos los requisitos que se detallan a continuación, se le concederá el certificado del POP.

- El colegio ha confirmado que el alumno ha completado los estudios de formación profesional específicos.
- Haber obtenido, como mínimo, un 3 en al menos dos de los cursos del Programa del Diploma.
- Haber obtenido, como mínimo, una D en el Proyecto de Reflexión.
- El colegio ha confirmado que se han cumplido todos los requisitos de Habilidades Personales y Profesionales, Aprendizaje-servicio y Desarrollo de la Lengua.
- No haber recibido una sanción del Comité de la evaluación final por conducta improcedente.
- Las calificaciones obtenidas en las asignaturas del PD y en el Proyecto de Reflexión las confirma el mismo Comité de la evaluación final que para el PD.

#### Certificado bilingüe del POP

En lugar del certificado del POP habitual, el alumno puede obtener un certificado bilingüe 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 debe haber obtenido como mínimo una calificación de 3 en ambos cursos.

#### Modo de calcular el resultado del PAI

Al final del quinto año del PAI, los alumnos pueden matricularse en la evaluación externa del IB. Los resultados de estas evaluaciones quedarán registrados en el documento de resultados de los cursos del PAI. Además, estos alumnos pueden decidir si quieren realizar las evaluaciones con las que se les podría conceder el certificado del PAI.

El colegio también puede emitir un título de participación del PAI, pensado para los alumnos del PAI que cursan el programa durante dos años como mínimo y cumplen con los requisitos correspondientes al tercer o cuarto año. Estos alumnos no están matriculados para ningún tipo de evaluación con el IB. El título de participación es un documento del colegio y no está verificado por el IB.

Para conseguir el certificado del PAI, el alumno debe participar en 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 dichas asignaturas
- Completar el examen en pantalla de la evaluación interdisciplinaria y obtener una calificación de 3, como mínimo



- Completar el Proyecto Personal con una calificación de 3, como mínimo
- Obtener una puntuación total de 28 puntos, como mínimo
- Cumplir las expectativas de servicio comunitario del colegio

El certificado bilingüe del PAI requiere además obtener resultados satisfactorios en las evaluaciones 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 que no sea la que el alumno eligiera para el curso de Lengua y Literatura

# Equidad para todos: cómo satisfacer las necesidades de los alumnos

Para que las evaluaciones del IB sean válidas, no pueden discriminar a los alumnos con necesidades especiales. Por ello, el IB tendrá en cuenta las peticiones de ofrecer exámenes modificados y adecuaciones inclusivas, tal como se establece en los *Procedimientos de evaluación* de cada programa.

- La mejor manera de garantizar la equidad es diseñar evaluaciones que sean accesibles para todos, lo
  cual eliminaría la necesidad de realizar modificaciones. Esta idea queda englobada en el concepto de
  diseño universal de evaluación que forma parte del compromiso del IB con el diseño universal para el
  aprendizaje (DUA).
- En ocasiones, las necesidades de algunos alumnos se conocen previamente y se les da respuesta con adecuaciones inclusivas de acceso, entre ellas el uso de cuestionarios de examen modificados.
- Sin embargo, a veces hay circunstancias que surgen en el último momento o que no se pueden solucionar mediante adecuaciones inclusivas. En dichos casos se aplican los procesos de consideración especial.
- En última instancia, el propósito de estas adecuaciones es que haya equidad para todos los alumnos. En consecuencia, a la hora de tomar decisiones, el IB debe tener en cuenta lo que es justo para todo el grupo y no solo para un alumno concreto. El objetivo es que todos los alumnos estén en igualdad de condiciones.

El IB cree que todos los alumnos deben tener la posibilidad de demostrar sus capacidades en condiciones de examen lo más equitativas posible. Las condiciones normales de evaluación pueden representar una desventaja para aquellos alumnos con necesidades de apoyo para el aprendizaje, impidiéndoles demostrar su nivel de conocimiento y capacidad. Asimismo, el IB es consciente de que, en ocasiones, hay circunstancias o hechos que escapan al control de los alumnos y que pueden afectar a su desempeño, por lo que se deberían tener en cuenta.

La mejor forma de garantizar la equidad en una evaluación es que todo el mundo realice la misma evaluación de la misma manera. Muchas de las modificaciones que se hacen para cubrir las necesidades específicas servirían para ayudar a todos los alumnos a entender y responder las preguntas. La situación ideal es que todas las evaluaciones se desarrollen teniendo en cuenta la variedad de necesidades que pueden tener los alumnos, en vez de tratar a algunos alumnos de manera distinta. En eso consiste el concepto de diseño universal de evaluación. El IB es consciente de que a veces no es posible adoptar este enfoque de inclusividad total, y por eso también cuenta con un proceso para solicitar adecuaciones inclusivas específicas.

Las **adecuaciones inclusivas de acceso** están diseñadas para satisfacer las necesidades específicas de los alumnos, por ejemplo, aquellos que tienen:

- Dificultades de aprendizaje
- Dificultades del lenguaje
- Dificultades de aprendizaje específicas
- Dificultades de comunicación o del habla
- Trastornos del espectro autista
- Dificultades sociales, emocionales o de conducta
- Discapacidades varias o trastornos motores o sensoriales, afecciones médicas o problemas de salud mental

Se considerará la realización de ajustes razonables para ciertos alumnos con el fin de satisfacer sus necesidades específicas. Para obtener más información, consulte los *Procedimientos de evaluación* del



programa correspondiente y la publicación del IB Alumnos con necesidades específicas de acceso a la evaluación.

**Circunstancias adversas** son todas aquellas circunstancias ajenas a la voluntad del alumno que puedan perjudicar su desempeño. El Comité de la evaluación final valorará dichas circunstancias y, en caso de decisión favorable, los alumnos que estén cerca de un límite de calificación recibirán la calificación superior. El principio de equidad para todos los alumnos, que defiende el IB, implica que, al considerar cualquier adecuación inclusiva o circunstancia adversa, no se debe crear una situación injusta para los otros alumnos que realicen las evaluaciones. El objetivo es que todos los alumnos estén en igualdad de condiciones.

## Principios para las adecuaciones inclusivas de acceso

Los principios para las adecuaciones inclusivas de acceso se establecen en la publicación del IB *Alumnos con necesidades específicas de acceso a la evaluación*. El texto que aparece a continuación se ha extraído de la versión del PD, pero en todos los programas se aplican principios similares.

- 1. El IB debe garantizar que la calificación final concedida a un alumno en cualquier asignatura no sea una descripción equívoca de su nivel de conocimiento y capacidad. Por tanto, se aplican los mismos estándares de evaluación para todos los alumnos, independientemente de si tienen o no necesidades de apoyo para el aprendizaje.
- 2. Las adecuaciones inclusivas de acceso, que incluyen ajustes razonables, son medidas que se adoptan antes de un examen para que un alumno pueda acceder a la evaluación. Dichas adecuaciones no pueden solicitarse con carácter retrospectivo, ya sea para un examen oral o escrito.
- 3. Las adecuaciones solicitadas para un alumno no deben representar una ventaja para dicho alumno en ningún componente de evaluación.
- 4. Las adecuaciones inclusivas de acceso descritas en este documento han sido concebidas para alumnos con un nivel de aptitud suficiente como para satisfacer los requisitos de evaluación necesarios para que se les otorque el diploma o los resultados de los cursos.
- 5. Cuando sea necesario, el colegio puede proporcionar adecuaciones inclusivas de acceso para un alumno, ya sea durante el curso de sus estudios en el Programa del Diploma o en exámenes de práctica. En este documento se enumeran las adecuaciones que pueden adoptarse para la evaluación sin que sea necesario solicitar la autorización del IB; en los demás casos, será obligatorio obtener autorización del centro global del IB en Cardiff. De igual modo, si un alumno del Programa del Diploma tiene problemas para satisfacer los requisitos de Creatividad, Actividad y Servicio (CAS), se deberá consultar escribiendo a "El IB responde".
- 6. Se recomienda a los colegios que planifiquen las adecuaciones inclusivas de acceso para sus alumnos basándose en los criterios del IB, tal como se establece en esta política y en las observaciones de los profesores del alumno en el aula durante las pruebas y el trabajo en clase.
- 7. Las adecuaciones inclusivas de acceso que sean solicitadas para un alumno deben ser las que tenga normalmente para trabajar durante el programa de estudios. El IB autorizará únicamente en casos muy poco comunes y realmente excepcionales aquellas solicitudes de adecuaciones inclusivas de acceso que no sean el modo normal de trabajo del alumno y que solo se hayan realizado para respaldar al alumno en algún momento de los seis meses anteriores a la petición, justo antes de los exámenes.
- 8. El IB tiene como propósito autorizar las adecuaciones inclusivas de acceso que sean compatibles con aquellas que normalmente están a disposición del alumno en cuestión. No obstante, solo se autorizarán las adecuaciones que sean coherentes con la normativa y las prácticas del IB. No se debe dar por supuesto que el IB admitirá las adecuaciones solicitadas por los colegios. Los coordinadores deben proporcionar información sobre el método habitual de trabajo del alumno en el aula.
- La filosofía educativa del IB se basa en la mentalidad internacional. Esto implica que su normativa de adecuaciones inclusivas de acceso puede no reflejar las prácticas habituales de un país en particular.

- Esta normativa es resultado de un estudio de las prácticas establecidas en diferentes países, con el objeto de lograr la equidad entre alumnos con necesidades específicas de acceso a la evaluación.
- 10. El IB se asegurará, siempre que sea posible, de que se aplican las mismas disposiciones para alumnos con similares necesidades específicas de acceso. Debido a las diferencias culturales que existen con respecto a la identificación de las necesidades de apoyo para el aprendizaje y a la naturaleza de las adecuaciones de acceso autorizadas en los colegios, un cierto grado de flexibilidad puede ser necesario para garantizar la equiparación entre alumnos de diferentes países.
- 11. 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 haya solicitado el coordinador.
- 12. El IB trata toda la información sobre los alumnos de forma confidencial. Solo podrán acceder a dicha información, de ser necesario, el personal habilitado del IB y los miembros del Comité de la evaluación final, a quienes se les pide que respeten su carácter confidencial.
- 13. Si un colegio no satisface las condiciones especificadas por el IB a la hora de administrar las adecuaciones inclusivas de acceso, o introduce disposiciones especiales sin autorización, es posible que al alumno no se le conceda una calificación final en la asignatura y nivel en cuestión.
- 14. Si se puede demostrar que la falta de dominio de la lengua de respuesta de un alumno es consecuencia de una necesidad de apoyo para el aprendizaje que ha sido identificada, se pueden autorizar adecuaciones inclusivas de acceso. (Para las asignaturas en los grupos 3 a 6, todos los alumnos pueden usar un diccionario bilingüe en los exámenes escritos).
- 15. Los colegios no deben informar a los examinadores de las circunstancias adversas o la situación particular de un alumno (como autismo, dificultades para escribir, etc.).
- 16. En el caso de los trabajos evaluados internamente, los profesores no deben hacer ningún ajuste al calificar el trabajo del alumno.
- 17. La lista de las adecuaciones inclusivas de acceso permitidas se revisa regularmente. El IB tomará en consideración adecuaciones alternativas propuestas por los coordinadores, siempre y cuando todos los alumnos con necesidades similares puedan beneficiarse de dichas adecuaciones.

De acuerdo con el documento *Reglamento general del Programa del Diploma*, los alumnos que cursan este programa pueden participar en tres convocatorias de exámenes para obtener el diploma. A criterio del IB, un alumno con necesidades de apoyo para el aprendizaje podrá participar en convocatorias adicionales.

- 1. Si la naturaleza de la dificultad de un alumno o de la adecuación inclusiva de acceso que se haya autorizado pudieran causar molestias a otros alumnos durante un examen, el alumno deberá realizar el examen en una sala independiente y ser supervisado de acuerdo con el reglamento que rige la realización de los exámenes del Programa del Diploma.
- 2. Los exámenes escritos deben ser supervisados de acuerdo con el reglamento que rige la realización de los exámenes del Programa del Diploma. El supervisor de examen no deberá ser un familiar del alumno ni ninguna persona con la que pueda existir un conflicto de intereses evidente o aparente.
- 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 durante los exámenes deberán comunicarse a "El IB responde" lo antes posible.

### Exenciones de evaluación

Normalmente no se autorizan exenciones respecto de los componentes de evaluación. No obstante, si un componente o una parte de la evaluación exige una función fisiológica que un alumno 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 todas las alternativas posibles. Solo se autorizará una exención cuando existan motivos importantes. La incapacidad física del alumno 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 el documento Alumnos con necesidades específicas de acceso a la evaluación del programa correspondiente.

## Oportunidades para adecuaciones inclusivas con la evaluación en pantalla

El uso de la evaluación en pantalla permite que el alumno tenga más control sobre el modo en que le gustaría que se presentara la evaluación. Gracias a las computadoras, las evaluaciones se pueden presentar en distintos tipos de fuente, tamaños de texto y colores para satisfacer las necesidades de los alumnos. Este tipo de ajustes se puede poner fácilmente a disposición de todos los alumnos.

Actualmente, en muchos casos la adecuación de inclusión que se solicita es permitir utilizar una computadora. Esta necesidad se ve claramente satisfecha con la evaluación electrónica, siempre y cuando el software requerido que permita la inclusión sea compatible con la herramienta en pantalla. El IB es muy consciente de este requisito y procura que todas las evaluaciones en pantalla cumplan con los estándares habituales en el sector en lo referido al software que posibilita la inclusión.

#### Circunstancias adversas

Circunstancias adversas o imprevistas son todas aquellas circunstancias aienas a la voluntad del alumno 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. 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 todos los alumnos cumplan 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, consulte los Procedimientos de evaluación y el Reglamento general del programa correspondiente.

En esos casos, el Comité de la evaluación final estudiará las pruebas que aporte el colegio, con objeto de determinar si el alumno afectado puede tener derecho a consideraciones especiales. Si las circunstancias se consideran "adversas" y dignas de "consideración", se podrá realizar un ajuste de la puntuación total del alumno para las asignaturas o los componentes troncales del programa en cuestión. Si el alumno se encuentra a uno o dos puntos del límite de calificación superior para la asignatura o las asignaturas en cuestión, la calificación se aumentará.

#### Diseño universal de evaluación

En los párrafos anteriores se ha explicado cómo gestiona el IB las situaciones en las que los alumnos necesitan modificaciones o ayuda específica para poder realizar de manera justa las evaluaciones. A pesar de todo, la mejor solución es que las evaluaciones no supongan ninguna barrera para los alumnos. El concepto de diseño universal de evaluación se refiere a plantearse el acceso, la inclusión, la igualdad, las sensibilidades culturales, los estereotipos y los sesgos desde el primer diseño de la evaluación. Esto incluye la creación de las tareas y las preguntas de examen, y además aborda el diseño del modelo general de evaluación que determina cómo crear evaluaciones equiparables para cada convocatoria de exámenes. Si se diseñan evaluaciones más inclusivas desde el principio, no solo aumentará la pertinencia del constructo, sino que también se minimiza el riesgo de no satisfacer las necesidades de todos los alumnos.

El diseño universal de evaluación es un aspecto del concepto más general de diseño universal para el aprendizaje (DUA). El DUA procura crear entornos de aprendizaje accesibles para todos los alumnos, incluidos aquellos con discapacidades, de diversas procedencias culturales y lingüísticas, superdotados y talentosos. Los principios del DUA se pueden aplicar en todas las áreas de la educación, en particular en el diseño del currículo, en la gestión escolar y en la enseñanza. Para obtener más información sobre el DUA en el contexto del IB, consulte: Rao, K.; Currie-Rubin, R. y Logli, C. *UDL and Inclusive Practices in IB Schools Worldwide*, 2016.

#### 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 a los alumnos 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 de los alumnos.

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 a la resolución de problemas de probidad académica.



## Consultas sobre los resultados, apelaciones y comentarios generales

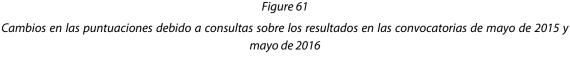
- El propósito de los servicios de consulta sobre los resultados es que los colegios puedan indicar al IB dónde creen que se ha producido un error durante el proceso de corrección.
- Durante las consultas sobre los resultados deben aplicarse exactamente los mismos estándares de corrección que durante la convocatoria de exámenes principal. Para asegurarse de que así sea, el IB utiliza los exámenes de control previstos 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 que solo escriban comentarios para explicar su corrección.
- El IB cuenta con un procedimiento formal de apelaciones que los colegios o los alumnos pueden emplear si consideran que los procesos no se han seguido correctamente.

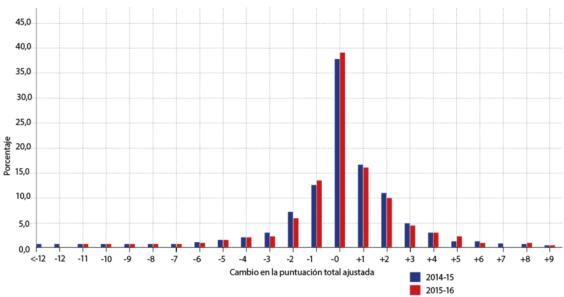
Para conocer la descripción de los aspectos legales y de los procedimientos de los servicios de consulta sobre los resultados, consulte los Procedimientos de evaluación del programa correspondiente.

## Propósito de la consulta 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.

Tal como se describe en la sección que trata sobre el modelo de calidad para la corrección, es posible que dos examinadores tengan un pequeño desacuerdo sobre cuántos puntos conceder a una respuesta concreta y, sin embargo, que ninguno de ellos esté equivocado. 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.





En el proceso de consulta sobre los resultados, el IB emplea únicamente a los examinadores más coherentes y con mayor nivel en la jerarquía. Habida cuenta de la jerarquía de los examinadores, se considera que el resultado de esta revisión de la corrección es la puntuación correcta.

Teniendo en cuenta la naturaleza del proceso para establecer los límites de calificación, siempre existe la posibilidad de que haya alumnos que se queden tan solo un 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, como se ha mencionado anteriormente, esto no implica necesariamente que haya habido algún error en la corrección. En este tipo de situación, el IB sostendría que el alumno se encuentra en el límite entre dos calificaciones, y que ambas reflejan de manera justa su desempeño.

A fin de minimizar los efectos de estas diferencias menores, aunque correctas, entre las puntuaciones de los examinadores, durante una consulta sobre los resultados se revisa la corrección de todos los exámenes externos del alumno. De este modo se mitiga el impacto de los pequeños cambios, ya que, siempre que no sean sistemáticos, se deberían compensar a lo largo de distintas pruebas.

#### Mantenimiento de los estándares

El propósito de las consultas sobre los resultados es proporcionar a los alumnos las puntuaciones que deberían haber recibido durante el período de corrección. Sin embargo, es muy importante que los examinadores no se dejen influir porque un alumno no esté satisfecho con su puntuación inicial, o porque dicha puntuación esté cerca de un límite de calificación.

Para que los examinadores mantengan el estándar de corrección adecuado, el IB incluye exámenes de control en el trabajo de consulta sobre los resultados. De este modo, los correctores reciben un aviso electrónico si se alejan del estándar establecido.

A fin de reducir las posibilidades de que un examinador cometa un error grave durante una consulta sobre los resultados, el IB revisa todos los cambios propuestos al terminar este procedimiento. Así, si entre los dos examinadores existe una diferencia muy notable que no se ha explicado, se pide una tercera opinión sobre

ese examen. En esos casos, normalmente se utiliza la puntuación que otorque el examinador con mayor nivel en la jerarquía.

## Identificación de problemas sistemáticos

El IB supervisa todos los cambios que se producen en las puntuaciones durante la consulta sobre los resultados, con el fin de detectar si hay algún patrón que indique algún problema con las puntuaciones originales. Este sería, por ejemplo, el caso de un examinador que haya superado los controles de calidad, pero que no haya seguido constantemente el mismo estándar de corrección.

Si hay pruebas de que se producen estos errores sistemáticos, el IB los investiga para determinar si se puede establecer una causa subyacente. En caso de que haya algún problema de este tipo, se identificará a todos los alumnos afectados (independientemente de si han solicitado una consulta sobre los resultados) y se comprobará que no se les haya perjudicado.

También es frecuente que los colegios comuniquen posibles casos de problemas sistemáticos. El IB se toma muy en serio estas inquietudes y estudiará todas las pruebas disponibles. Sin embargo, salvo excepciones, solo se realizará una revisión de la corrección si el colegio lo solicita mediante el proceso de consulta sobre los resultados.

## Categorías de consultas sobre los resultados

Tras la publicación de los resultados, el coordinador puede solicitar una consulta sobre los resultados de:

- Categoría 1: revisión de la corrección de todos los componentes de evaluación externa de un alumno en una asignatura
- Informe de categoría 1: informe sobre la revisión de la corrección de la consulta sobre los resultados de categoría 1
- Categoría 2: copias de material de algún componente de evaluación externa
- Categoría 3: revisión de la moderación de un componente de evaluación interna

Se debe abonar una tasa por cada una de las categorías anteriores, salvo cuando se modifica una calificación final como consecuencia de una revisión de la corrección de categoría 1.

Un examen devuelto puede contener comentarios útiles del examinador al que se le había asignado. Sin embargo, no se puede garantizar que sea así, ya que los examinadores no están obligados a escribir comentarios cuando corrigen el trabajo de los alumnos. Esto se debe a que el propósito de la evaluación sumativa es reflejar de manera fiel el desempeño del alumno al final de sus estudios, y lo importante es que los examinadores corrijan de manera adecuada, en lugar de aportar anotaciones y recomendaciones sobre el trabajo. Los comentarios sobre el desempeño del alumno conforman una parte importante de la evaluación formativa que realiza el profesor a lo largo del curso.

Tal como se comenta en la sección "Moderación" relativa al trabajo de evaluación externa, la puntuación que otorga un examinador en una prueba no necesariamente será la puntuación "moderada" final; la puntuación que se da al alumno en el sistema de información del IB (IBIS) es la correcta.

## Equidad para todos (aumento y disminución de las calificaciones finales)

- Tras una consulta sobre los resultados de categoría 1, las puntuaciones de los alumnos pueden aumentar o disminuir. Por ello, es obligatorio pedir a los alumnos permiso para que su trabajo pase por este procedimiento.
- No sería razonable que el IB exigiera a los colegios obtener permiso de todos los alumnos de una clase antes de solicitar una consulta sobre los resultados de categoría 3 (revisión de la moderación). Por lo tanto, este es el único caso en el que las calificaciones de los alumnos están protegidas y no pueden disminuir.

En los casos en que la puntuación de un alumno se vaya a cambiar por otro motivo, se estudiará caso
por caso si se debe proteger la calificación. Sin embargo, si la situación se debe a motivos ajenos a la
voluntad del alumno o del colegio, en principio no se permitirá reducir la calificación.

El sistema de consulta sobre los resultados pretende garantizar que los alumnos reciban la puntuación correcta para su trabajo, así que el IB debe prestar mucha atención para que la puntuación obtenida de una consulta sobre los resultados sea lo más precisa posible. Para ello, se pide a los examinadores más coherentes y con mayor nivel en la jerarquía que se encarguen de la consulta sobre los resultados y que revisen sus resultados para cerciorarse de que pueden justificar su decisión.

Con estas medidas de precaución, se considera que la puntuación que se decide tras una consulta sobre los resultados es la representación más fiel del trabajo del alumno, por lo que esa será la puntuación que se otorgará y que servirá para calcular la calificación final. Esto significa que, tras una consulta sobre los resultados de categoría 1, las puntuaciones de los alumnos pueden aumentar o disminuir. Por ello, es obligatorio que los colegios pidan a los alumnos permiso para que su trabajo pase por este procedimiento.

En lo que respecta a las consultas sobre los resultados de categoría 3 (revisión de la moderación), cualquier cambio en la puntuación podría afectar a todo el grupo de alumnos. Si en estos casos se aplicara también el principio de contar con el permiso de los alumnos, esto implicaría que si un solo alumno 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 de los alumnos sin que hayan aceptado formalmente este riesgo, las consultas sobre los resultados de categoría 3 solo pueden aumentar las calificaciones, y nunca disminuirlas.

Si la puntuación de un alumno cambia por cualquier otro motivo, se estudiará caso por caso si se debe proteger la calificación. Sin embargo, si la situación se debe a motivos ajenos a la voluntad del alumno o del colegio, entonces todo permite suponer que la calificación se protegerá, ya que puede que el alumno haya tomado decisiones irrevocables basándose en ese resultado.

# Comentarios de las evaluaciones corregidas

Las evaluaciones que se tratan en esta sección son evaluaciones sumativas. Aunque nada impide que se utilicen también con fines formativos, el IB debe procurar garantizar que su validez no se vea socavada por otros propósitos.

Se espera que los examinadores corrijan con precisión y de manera adecuada, respetando el estándar que establezca el examinador principal. Esto significa que solamente se les pide que incluyan comentarios para explicar su corrección, y no aportar comentarios a los alumnos para indicarles cómo podrían mejorar. Como saben los docentes, ofrecer comentarios instructivos implica destrezas distintas a las que se utilizan en la corrección sumativa, y, en principio, no es posible hacerlo adecuadamente sin tener un conocimiento holístico del alumno. Esto supone una contradicción directa con la expectativa del IB de que, al corregir, los examinadores no se vean influidos por absolutamente ningún sesgo sobre los alumnos.

Con tan solo reflexionar sobre lo que se ha hecho bien en la evaluación se puede ayudar a los alumnos y a los profesores a mejorar sus habilidades y comprensión. Por este motivo, el IB aporta con mucho gusto información sobre los resultados de la corrección, aunque partiendo de la premisa de que estos comentarios no son de carácter formativo.

## **Apelación**

El procedimiento de consulta sobre los resultados debe garantizar que se resuelvan todos los errores que haya en la corrección o en los procesos de evaluación. Si un colegio o un alumno siguen considerando que sus evaluaciones no se han tratado de manera adecuada, pueden presentar una apelación formal. Sin embargo, solo es posible optar por la apelación cuando se han agotado las demás alternativas.

Pueden presentarse apelaciones contra:

1. Resultados: si un colegio tiene razones para creer que los resultados de un alumno matriculado son erróneos después de agotar todos los procedimientos de consulta correspondientes



- Una decisión que confirma una conducta improcedente (sin embargo, no podrá presentarse una apelación contra la severidad de una penalización)
- Una decisión referente a una consideración especial, en caso de que esta se deniegue a un alumno matriculado que se vea afectado por presuntas circunstancias adversas
- Una decisión administrativa que no esté prevista en ninguno de los casos anteriores y afecte a los resultados obtenidos por uno o más alumnos matriculados

El proceso de apelación consta de dos etapas. En la primera etapa, un miembro sénior del personal del IB, que no haya participado en el caso hasta entonces, revisará las pruebas y la decisión. Si el apelante sigue sin quedar satisfecho, se pasará a la segunda etapa, que implica una revisión por parte de una comisión compuesta por tres personas, una de las cuales será externa al IB.

En el Reglamento general del programa correspondiente se encuentra información detallada sobre el proceso de apelación. Para obtener más información, escriba a appeals@ibo.org o comuníquese con"El IB responde".

Asimismo, el IB dispone de un ombudsman que actúa como mediador para "trabajar con personas y grupos de una organización con el objeto de ayudarlos a determinar y explorar opciones para resolver conflictos, situaciones problemáticas o inquietudes". Esta vía alternativa queda disponible para los colegios o los alumnos que no estén satisfechos con el proceso de evaluación.

# Preparación de las evaluaciones del año siguiente

- La evaluación se considera como un ciclo. Por tanto, es importante aprender de cada convocatoria de exámenes para mejorar continuamente las evaluaciones.
- Los cuestionarios de examen y los esquemas de calificación se preparan hasta dos años antes de la convocatoria, de manera que se tienen en cuenta distintas cuestiones en diferentes momentos durante el ciclo de vida de la evaluación.
- Las pruebas que se deben tener en cuenta van desde las experiencias operativas hasta los resultados de los alumnos, pasando por los comentarios de los profesores.
- Un principio importante para el IB es que se utilicen todas las herramientas disponibles para comprender y corregir cualquier problema relacionado con la equiparabilidad entre asignaturas, y no solamente para realizar ajustes a los límites de calificación.
- Cualquier cambio que pueda afectar a la pedagogía deberá implementarse con una antelación equivalente a la duración del curso (es decir, dos años).

# La utilización de las pruebas

La autorreflexión es un aspecto esencial de la filosofía del IB, y es tan importante en el proceso de evaluación como para los alumnos a los que se evalúan. Cada convocatoria de exámenes brinda la oportunidad de obtener una variedad de pruebas que muestren:

- Cómo interpretaron los alumnos las preguntas de examen y en qué medida lo hicieron en consonancia con las expectativas del IB
- Si las distintas preguntas cumplieron las expectativas previstas: ¿se sopesaron correctamente la exigencia y la dificultad?
- Con qué grado de fiabilidad lograron corregir cada pregunta los examinadores: ¿hubo alguna pregunta en la que fuera especialmente difícil mantener el estándar correcto?
- Si los distintos componentes cumplieron las expectativas previstas: ¿los porcentajes del total de la evaluación coincidían con los que se habían previsto?
- Si hubo problemas en la preparación o la corrección de los exámenes
- Si los profesores consideraron que el examen reflejaba sus expectativas: ¿hubo algún un problema a este respecto en alguna pregunta concreta?
- Si a los profesores les pareció que el examen era más fácil o más difícil que en convocatorias anteriores: ¿por qué?

Esta información se analiza y se utiliza de varias maneras dependiendo de la naturaleza de la cuestión. Si se trata de problemas con las preguntas que se están preparando, normalmente se adapta el estilo durante la fase de redacción de los exámenes (hasta dos años antes de que los realicen los alumnos), a no ser que dichos problemas sean tan graves que haya que reescribir los exámenes que ya estén preparados.

Por el contrario, en los casos en que las pruebas indiquen que se deben ajustar los procesos de evaluación o los estándares de corrección, esto puede hacerse para la siguiente convocatoria.

# Equiparabilidad entre asignaturas

La equiparabilidad es uno de los aspectos más complejos de la validez. Tal como se ve en la siguiente figura, a menudo no queda claro cuándo resulta razonable hacer comparaciones entre dos cosas distintas.



Comparar asignaturas también es complejo porque, por ejemplo, ¿cómo se compara el desempeño en la asignatura de Literatura con el desempeño en Matemáticas?

Figura 62 ¿Qué elementos se pueden comparar de manera razonable? ¿Se utiliza un enfoque relativo, o absoluto?



La situación se complica aún más porque los alumnos son mejores en unas asignaturas que en otras. Por lo tanto, no existe un alumno "promedio" que se pueda tomar como referencia para determinar la dificultad de una evaluación. Se han realizado muchas investigaciones sobre cómo medir la equiparabilidad entre asignaturas. Un ejemplo es el estudio de Ofqual (2015) titulado *Inter-Subject Comparability: A Review of the Technical Literature: ISC Working Paper 2*.

En los programas del IB, normalmente los alumnos deben estudiar un currículo amplio, compuesto por asignaturas que deben elegir de distintos grupos. Por tanto, es fundamental velar para que las asignaturas que forman esos grupos tengan el mismo nivel de exigencia, con el fin de que todos los itinerarios de un programa supongan una dificultad equivalente. Esto significa, por ejemplo, procurar que Biología y Física, o Geografía e Historia, sean equiparables.

También es importante hacer comparaciones entre los grupos de asignaturas, con el fin de que todos los alumnos reciban un tratamiento igualitario, independientemente de cuáles sean sus puntos fuertes.

Por encima de todo, el IB busca un equilibrio entre las distintas formas de aportar pruebas de desempeño, para que ninguna tenga más peso que las demás en la toma de decisiones.

En algunas materias, especialmente las de lengua, es posible que los examinadores cotejen trabajos de distintas asignaturas para comparar el nivel de dificultad. Normalmente esto se hace durante el proceso de estandarización.

Después de cada convocatoria de exámenes, el IB realiza un análisis entre pares de asignaturas. Este método bien documentado (Nuttall et al., 1974), consiste en observar a alumnos que hayan cursado las dos asignaturas en cuestión y comparar sus calificaciones finales en ambas. Así, en cada asignatura se puede calcular una diferencia media en relación a las demás asignaturas y establecer una clasificación según la dificultad. Aunque este método presenta limitaciones (véase Ofqual, 2015 o Coe et al., 2008), si se combina con otras pruebas constituye una fuente de información muy valiosa.

Si el análisis entre pares de asignaturas o las opiniones de los expertos en las asignaturas indican que hay alguna diferencia de nivel entre las asignaturas, esto se discute durante el proceso de concesión de calificaciones finales. Por ejemplo, si a raíz de realizar un análisis entre pares de asignaturas se detecta que una asignatura es "difícil", el IB esperaría que los examinadores se mostrasen generosos al realizar sus

valoraciones. Un principio importante para el IB es que se utilicen todas las herramientas disponibles para comprender y corregir cualquier problema relacionado con la equiparabilidad entre asignaturas, y no solamente para realizar ajustes a los límites de calificación.

## Realización de cambios

Al hacer cambios en las evaluaciones, se parte de un principio importante: no perjudicar ni a los alumnos ni a los profesores. Esto significa que, si una modificación puede afectar a las prácticas de enseñanza o a la preparación para los exámenes, solamente se aplicará para aquellos alumnos que estén a punto de empezar el curso.

# Apoyo a las revisiones del currículo

Las pruebas obtenidas sobre el desempeño en una evaluación se utilizan no solo para mejorar las futuras evaluaciones, sino también para facilitar el proceso de revisión del currículo. En todas las revisiones del currículo se presenta un informe formal sobre la evaluación. Además, por lo general, los responsables curriculares del IB participan en la primera reunión de concesión de calificaciones de cada curso, con el fin de observar directamente las experiencias de los examinadores.

# Evaluaciones en pantalla

La evaluación en pantalla no debería ser distinta a la tradicional en lo referido a aprender lecciones de convocatorias anteriores. La evaluación en pantalla permite plantear preguntas innovadoras y originales, lo cual es una gran ventaja. Sin embargo, dichas preguntas se deberán inspeccionar minuciosamente con el fin de que satisfagan las expectativas.



# Comentarios para los colegios

- El propósito de la evaluación sumativa del IB es medir el desempeño de los alumnos, y todos los procesos deben diseñarse de tal modo que permitan obtener los resultados más válidos que sea posible.
- La transparencia forma parte de esto, con el fin de que los alumnos y los profesores vean que las puntuaciones y las calificaciones se otorgan de manera coherente y sin sesgos.
- Este aspecto es de especial importancia en la evaluación interna, ya que la calificación final se basa en una moderación de las puntuaciones del profesor.
- 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 de los alumnos, 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.

## Comentarios de los examinadores

No es posible ocuparse de dos cosas a la vez. Es como las calabazas en el agua: una sale a la superficie mientras se intenta mantener la otra hundida.

Proverbio chino.

El propósito de la evaluación sumativa del IB es medir el desempeño de los alumnos. Por tanto, se pide a los examinadores que se centren únicamente en corregir el trabajo de los alumnos según el estándar establecido. Los examinadores solo deben incluir comentarios si esto les ayuda en su labor de corrección.

Para poder ofrecer comentarios formativos, ya sean para los alumnos o para los profesores, el examinador debe determinar la puntuación correcta que corresponda al trabajo del alumno y, a continuación, intentar explicar cómo se podría mejorar dicho trabajo. 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 de su cometido principal, que es corregir según un estándar coherente. En términos sencillos, el IB quiere que los examinadores realicen una tarea (corregir) con la máxima calidad, y no dos tareas (corregir y proporcionar comentarios) con una calidad menor.

El examinador solamente puede ofrecer su valoración a partir del trabajo que tiene a su disposición, mientras que los profesores con experiencia pueden utilizar una amplia variedad de información para decidir qué comentarios aportar a un alumno. Esto significa que los comentarios del examinador tendrán una perspectiva más limitada que los del profesor.

Por todos estos motivos, el IB deja muy claro a los examinadores que deben limitarse a corregir el trabajo de los alumnos siguiendo el estándar establecido, y que no deben añadir comentarios para el alumno ni para el profesor.

Los examinadores 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 puntos sea más transparente para los colegios.

El informe de categoría 1, que los colegios pueden solicitar en el contexto de una consulta sobre los resultados, constituye la única excepción a este principio de proporcionar comentarios solo para explicar la corrección. En ese caso, un examinador supervisor responderá las inquietudes concretas que haya señalado el colegio al solicitar la consulta sobre los resultados, y aportará comentarios más detallados que los que normalmente se esperan de los examinadores.

# Informes generales de las asignaturas

Todos los examinadores jefe tienen que generar un informe general de la asignatura después de cada convocatoria. El propósito de este informe es dar a los profesores información sobre el desempeño del grupo de alumnos que han participado en una convocatoria de exámenes concreta, y en él se tratan las preguntas y los temas con mejores y peores resultados. En el informe se aportan datos sobre cada componente, se discute la calidad general de las respuestas de los alumnos y se ofrecen recomendaciones generales sobre cómo se podrían mejorar las respuestas en futuras convocatorias. Asimismo, muchos informes de las asignaturas proporcionan datos clave basados en los comentarios que los profesores hicieron acerca del examen y que se tuvieron en cuenta durante la concesión de calificaciones finales. Finalmente, en estos informes se proporcionan los límites de calificación para la asignatura, así como los límites de los componentes que se utilizaron para alcanzar los resultados totales.

Lógicamente, cada profesor tendrá que situar toda esta información en el contexto de su propia clase. Por ejemplo, aunque una pregunta concreta haya obtenido respuestas mediocres en general, es posible que todos los alumnos de una clase hayan obtenido puntuaciones altas en dicha pregunta.

## Comentarios sobre la evaluación interna

El propósito de la moderación de la evaluación interna es garantizar que todos los profesores siguen el mismo estándar de corrección. En última instancia, al IB le gustaría que no se tuviera que aplicar ningún factor de moderación a los trabajos de los colegios. Para ayudar a los docentes 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 permite a los profesores 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 los alumnos de la muestra podrían haber conseguido mejores resultados.

# 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 el IB, pretenden servir de ayuda para que los profesores y los alumnos 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 todos los profesores 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 a los profesores a interpretar mejor dicho contenido. Si se realizan modificaciones a las guías, el IB publicará formalmente estos cambios y los describirá claramente como tales.



# ¿En qué consisten los procesos específicos de los programas?

- El propósito general de contar con evaluaciones sumativas externas en los programas del IB es proporcionar a los alumnos un "pasaporte" que les ayude a avanzar en su educación o en el mundo laboral.
- Puesto que todos los programas del 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 los alumnos avancen por los distintos programas en las diferentes etapas de la maduración educativa.
- Si bien la evaluación es un elemento importante del 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.

# Propósito de contar con evaluaciones sumativas externas en los programas del IB

El propósito de tener una evaluación sumativa externa en el PD, el POP y el PAI (en este caso, opcional en el último año) es proporcionar a los alumnos un "pasaporte" que les ayude a avanzar en su educación o en el mundo laboral.

El IB debe organizar esta evaluación sumativa externa de tal modo que el "pasaporte" que se le ofrece los alumnos cumpla con la declaración de principios y la filosofía del IB, cuyos principales elementos se resumen a continuación.

- Dar prioridad a una educación holística basada en programas, y no a asignaturas limitadas
- Valorar la mentalidad internacional, mediante el perfil de la comunidad de aprendizaje
- Valorar las oportunidades de evaluación auténtica y significativa por encima de la evaluación procedimental basada en recordar la información rápidamente
- Permitir una diferenciación justa de los alumnos en los contextos adecuados
- Fomentar un efecto de repercusión positivo en la enseñanza y el aprendizaje

Todos estos valores están reflejados en los principios de evaluación del IB.

Los alumnos que cursan el PEP necesitan mucho menos ese "pasaporte", lo cual explica la ausencia de una evaluación sumativa externa. Sin embargo, se aplican de todos modos los principios de buena evaluación en los que se basa la filosofía del IB.

El propósito secundario de la evaluación en el IB es dar ejemplo en un mundo en el que se valora mucho la evaluación.

Por último, la evaluación sumativa sirve para que profesores, colegios, padres y alumnos tengan la certeza de que el concepto de educación de alta calidad del colegio coincide con el del IB.

# Necesidades y soluciones específicas de los programas

En las secciones anteriores se han explicado los procesos que conforman el ciclo de evaluación, y 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. Como ejemplos pueden mencionarse cómo se formulan y se evalúan la Monografía o el Proyecto Personal, o cuáles son las reglas que determinan si se concede el título 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.

# Transición del diseño curricular a la evaluación y vuelta al diseño curricular

En las dos secciones anteriores se ha explicado que la evaluación debe estar integrada en los objetivos generales del curso. En consecuencia, el modelo de evaluación debe formar parte del desarrollo del curso e influir en dicho desarrollo para que los resultados de la evaluación reflejen de manera justa y significativa aquello que quiere proporcionar el curso.

En el IB, el enfoque de la revisión curricular de las asignaturas incluye la discusión sobre la evaluación. De hecho, entre los materiales para profesores que se ofrecen a los colegios figuran ejemplos de cómo serán las futuras evaluaciones, y también muestras de trabajos de alumnos con anotaciones para explicar la corrección o la evaluación.

Los ejemplos de material de examen, como los exámenes de muestra, tienen los siguientes propósitos:

- Mostrar la estructura de los exámenes, es decir, en qué partes los alumnos deben responder todas las preguntas de una sección, en qué partes hay preguntas opcionales, y la extensión y los puntos disponibles de cada sección y de cada prueba
- Indicar el estilo de las preguntas y de los tipos de material de estímulo que habrá en el examen, particularmente para los elementos nuevos de un curso
- Proporcionar una idea del contenido de un examen real, sobre todo en lo relativo a los elementos del resumen del programa de estudios (o del currículo) que son obligatorios y los que son opcionales

Proporcionar a los profesores material para preparar **de manera razonable** a los alumnos para los exámenes, por ejemplo, mediante simulacros de examen

Los exámenes de muestra pueden contener ejemplos de preguntas que traten acerca de todos los aspectos del curso, incluso los aspectos nuevos. Los exámenes de muestra se deben diseñar de tal modo que se pudieran utilizar eficazmente como exámenes "reales" del IB, y no solo como material ilustrativo.



# 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 la enseñanza y el aprendizaje.
- Estos son los elementos comunes que conectan a todos los programas del IB:
  - El perfil de la comunidad de aprendizaje
  - Los enfoques la enseñanza y el aprendizaje
  - La mentalidad internacional



Figura 63

En la publicación ¿Qué es la educación del IB? 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. En los programas se destaca 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 a los alumnos 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 a los alumnos 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. Los alumnos entran
en contacto con muchas disciplinas simultáneamente a lo largo de sus programas de estudio;
aprenden a establecer conexiones y enriquecen 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.

En ese contexto, la evaluación de todos los programas incluye un proyecto culminante. En el PEP se trata de la exposición, en el PAI es el Proyecto Personal o el Proyecto Comunitario, en el PD es la Monografía, y en el POP es el Proyecto de Reflexión.

Incluso cuando diseñen evaluaciones individuales, los profesores, los colegios y los autores del IB tienen que reflexionar sobre estos objetivos subyacentes para no crear tareas que obstaculicen una buena enseñanza y un buen aprendizaje. El propósito siempre debe ser generar un efecto de repercusión positivo.

Para cumplir con la declaración de principios del IB, los programas de buena calidad deben incluir tres áreas:

- 1. El perfil de la comunidad de aprendizaje
- 2. Enfoques de la enseñanza y enfoques del aprendizaje
- 3. Mentalidad internacional y entendimiento intercultural

Para que las evaluaciones del IB cumplan con los objetivos generales de los programas y resulten adecuadas al propósito (válidas), también deben considerar estos elementos, aunque no sean lo que se quiere evaluar.



# Las competencias del alumno y 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.

Cada vez se afirma más que este siglo requiere habilidades radicalmente distintas a las que se esperaban en las generaciones anteriores. Si bien hay quienes argumentan que el enfoque de indagación que subyace a las habilidades del siglo XXI se ha valorado desde Sócrates, existe cierto consenso general sobre la importancia de proporcionar a los alumnos una amplia variedad de atributos para prepararlos para la vida. Véanse, por ejemplo, los argumentos que se dan en Llewellyn *et al.*, 2013.

Hay muchas maneras distintas de categorizar estas habilidades, por ejemplo, el marco de competencias para el siglo XXI de la OCDE, el que propone RAND Education, y el del National Research Council (NRC). El IB, por su parte, describe estas competencias dentro del perfil de la comunidad de aprendizaje.

ORES DO SE INSTRUIDOS SOLIDA BUENOS COMUNICADORES DE MENTALIDADOS E INSTRUIDOS SOLIDA BUENOS COMUNICADORES DE MENTALIDADOS COM

Figura 29
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, 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, aunque no esté diseñada para medirlas, puede ofrecer a los alumnos 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 e introducir contextos inesperados a los alumnos.

Si quiere obtener más información sobre el enfoque amplio del IB respecto a las competencias de los alumnos, consulte el material disponible en el sitio web del IB o el documento *De los principios a la práctica* del programa pertinente.

# Enfoques de la enseñanza y enfoques del aprendizaje

Con sus enfoques del aprendizaje y sus enfoques de la enseñanza, el IB no pretende imponer prescripciones o restricciones, sino ofrecer consejos y sugerencias. No cabe duda de que los profesores y los colegios deben disponer de suficiente libertad para ser creativos, pero igualmente es necesario orientar las discusiones y destacar las buenas prácticas. El IB considera que, para conseguir que la enseñanza de las habilidades sea eficaz, los enfoques del aprendizaje deben definirse explícitamente e implementarse de manera continua.

Para mejorar una habilidad, es necesario reforzarla a lo largo de un período prolongado y en una variedad de contextos. Sea cual sea el método que los profesores o los colegios decidan utilizar para incluir los enfoques del aprendizaje en sus clases, el IB considera que se trata de una decisión que deben tomar aquellos que mejor conocen las necesidades de los alumnos, es decir, sus profesores.

Para obtener más información sobre este tema, consulte los recursos dedicados a los enfoques de la enseñanza y el aprendizaje, o las secciones correspondientes de la publicación *De los principios a la práctica* del programa pertinente.

## Mentalidad internacional y entendimiento intercultural

Los alumnos que estudian los programas del IB se encuentran en numerosos países y tienen una amplia variedad de nacionalidades. Además de los objetivos académicos de nuestros programas, el IB pretende que los alumnos se formen como "jóvenes solidarios capaces de contribuir a crear un mundo mejor y más pacífico, en el marco del entendimiento mutuo y el respeto intercultural" y alentarlos a "entender que otras personas, con sus diferencias, también pueden estar en lo cierto" (declaración de principios del IB, 2002). Por lo tanto, en la enseñanza el IB hay un contexto internacional y un propósito de entendimiento intercultural, y ambos elementos deben reflejarse en la evaluación.

Lo más importante para lograr esto es emplear a expertos académicos —entre otros, los autores de los cuestionarios de examen y los encargados del desarrollo curricular— procedentes de distintos contextos culturales. Para cumplir los principios del IB es importante no ocultar las diferencias, sino utilizarlas de manera que los alumnos puedan explorarlas sin verse desfavorecidos.

En algunas áreas disciplinarias, para fomentar la variedad cultural se pueden tener en cuenta distintas perspectivas culturales en el currículo. Podemos encontrar ejemplos de este enfoque en Biología, Química, Psicología y Artes Visuales. En las tres primeras, las estructuras de opciones 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, la mentalidad internacional se potencia a través de los materiales y las fuentes de inspiración que se fomenta que use el alumno. Algunos ejemplos en este sentido son las asignaturas de Artes, Literatura y Lengua, pero esta noción también se puede incluir en una amplia variedad 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, ya que esta se centra en los logros más que en los atributos afectivos.

En el IB esto se aborda con los elementos que no se evalúan de sus programas, como Creatividad, Actividad y Servicio (CAS) en el PD, y el Proyecto Comunitario en el PAI. Como los alumnos solo pueden obtener el título del correspondiente programa si completan los elementos no evaluados, estos tienen un impacto significativo en el resultado global de la evaluación del IB.

Figura 30
Acción basada en principios

Tal vez se considere que permitir a los alumnos elegir qué preguntas quieren responder es la mejor manera de abordar los distintos requisitos internacionales en materia de evaluación, pero esto provoca problemas en lo que se refiere a mantener la equiparabilidad entre las opciones. Esto siempre sucede cuando se pueden elegir las preguntas, o cuando las tareas de evaluación son muy abiertas. Resulta complicado siquiera definir en qué consiste una "exigencia igual" cuando los alumnos tienen bagajes educativos muy distintos. En general, la manera más fácil de mantener la equiparabilidad es establecer tareas comunes que

permitan a los alumnos apoyarse en sus propias experiencias para dar las respuestas. En esos casos, el examinador afronta el reto de mantener un estándar común, si bien esto resulta un poco más fácil que corregir según el mismo estándar dos tareas que pueden tener niveles de exigencia diferentes.

Para obtener información sobre la equiparabilidad, se puede analizar el desempeño de los alumnos. 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. Las preguntas que se refieren a deportes, viajes, entretenimiento, acontecimientos históricos, incluso al tiempo meteorológico, deben prepararse con sumo cuidado. Puede parecer que el único modo de evitar este problema sea preparar preguntas de examen carentes de todo contexto sociocultural excepto el mínimo común denominador. Sin embargo, hacerlo así no solo limitaría mucho las preguntas de exámenes y las haría aburridas, sino que iría en contra de toda la filosofía de evaluación del IB y contra el correcto ejercicio de la actividad de evaluación en lo concerniente a garantizar su validez mediante tareas basadas en contextos. La evaluación y el trabajo contextualizados son esenciales para el buen aprendizaje.

Existen dos posibles soluciones para este dilema. En primer lugar, puede proporcionarse a los alumnos información contextual complementaria. 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 la misma pregunta de examen (siempre que no ocupe demasiada extensión, con el fin de evitar que distraiga a los alumnos de la finalidad de la evaluación).

Un segundo método consiste en utilizar preguntas y tareas de evaluación más abiertas que permitan a los alumnos elegir el contexto en el que responder. En este enfoque, la puntuación ha de centrarse en niveles de comprensión más profundos, y no en conocimientos directos del contenido de la asignatura, ya que no habrá base común de contenido. Esto está muy en la línea de la filosofía de evaluación del IB.





Figura 31 Variedad de normas y contextos culturales

Incluso aplicando estos dos métodos, los alumnos pueden encontrarse con tareas de evaluación cuyos contextos no les resulten familiares en su propio medio sociocultural. También esto está en la línea de la filosofía de evaluación del IB, dado que uno de los objetivos de los programas es abrir las mentes de los alumnos a otras formas de hacer las cosas, aumentar su conciencia global y su competencia para operar en un entorno cultural que no les resulte familiar. Parte del requisito del pensamiento abstracto es que los alumnos deben poder aplicar sus conocimientos en situaciones desconocidas. Es apropiado incluir tales elementos en la evaluación, siempre que afecten por igual a alumnos de diferentes culturas.

Un número significativo de alumnos del IB se presentan a exámenes en una lengua que no es aquella en la que son más competentes. En casi todos estos casos se trata del inglés, porque los alumnos que estudian en francés o español (las otras dos lenguas en las que se realiza la evaluación) suelen ser hablantes nativos. Debe procurarse de manera especial que la redacción de las preguntas no sitúe en desventaja a los alumnos que realizan la evaluación en su segunda lengua. De esto se encargan los responsables de la edición de exámenes.

La evaluación sumativa del IB, al igual que la gran mayoría de sistemas de evaluación formales, es muy individualista. Como señaló Brown (2002), esto se debe en gran medida a que el Programa del Diploma pertenece a la tradición europea occidental, y las sociedades europeas occidentales son individualistas por naturaleza. Se evalúa a los alumnos casi exclusivamente sobre lo que logran por sí mismos. Puede decirse que culturalmente esto carece de equidad, ya que hay algunas culturas en las que lo que aporta el individuo está siempre subordinado a lo que aporta un grupo más amplio: lo que importa es lo que logra el grupo. También sucede que, en términos de equidad individual, hay algunas personas que trabajan mejor en equipo que individualmente, y viceversa. Además, es práctica habitual que, tanto en el aula como en el mundo laboral, los individuos trabajen de forma interdependiente más que independiente.

## 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?
- Enfoques de la enseñanza y el aprendizaje
- Perfil de la comunidad de aprendizaje del IB
- La publicación De los principios a la práctica correspondiente a cada programa (el PAI, el PD y el POP)



## El Programa del Diploma del IB

Las características distintivas de la evaluación sumativa externa del PD son:

- Los alumnos deben 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. Hay grandes variaciones en lo que respecta al
  número de alumnos que cursan cada asignatura.

## Objetivos generales del Programa del Diploma

La validez de los resultados de la evaluación solo se puede determinar si se tiene claro el propósito del curso y del programa. Por tanto, esta sección comienza con una presentación de los objetivos generales del programa.

El Programa del Diploma ofrece a los alumnos 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. Está concebido para desarrollar las habilidades académicas básicas que requerirán sus estudios universitarios, sus estudios posteriores y la profesión que elijan. Además, el programa apoya el desarrollo de las habilidades prácticas y los valores para una vida plena y fructífera.

El Programa del Diploma: de los principios a la práctica, 2015.

Lo que hace que el Programa del Diploma (y todos los demás programas del IB) sea especial es el interés que se pone en la experiencia educativa integral de cada alumno. Esto queda reflejado en el hincapié que se hace en la validez del programa en su conjunto, y no solo de cada curso individual.

#### 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 la entrada en la universidad o en el mundo laboral
- Si los alumnos ya han cumplido algunos requisitos de algún programa universitario (si se les pueden otorgar créditos o validar alguna asignatura o curso)

En los casos en los que el alumno ha realizado la evaluación en una lengua (de respuesta) concreta, eso también demuestra que puede cursar estudios posteriores de esa materia en dicha lengua.

Para obtener más información sobre la importancia de estos usos válidos, consulte la sección "¿Qué se considera válido: la evaluación o su uso?".

## La estructura del PD

Para obtener el diploma del IB, los alumnos deben cursar seis asignaturas y también los componentes troncales: Teoría del Conocimiento, la Monografía, y Creatividad, Actividad y Servicio (CAS).

Los alumnos eligen cursos de los siguientes grupos de asignaturas:

- Estudios de Lengua y Literatura
- Adquisición de Lenguas
- Individuos y Sociedades
- Ciencias
- Matemáticas
- Artes

Los alumnos pueden optar por estudiar un curso adicional de Ciencias, Individuos y Sociedades, o lenguas, en lugar de un curso de Artes.

Los alumnos deben cursar algunas asignaturas de Nivel Medio (NM) y otras de Nivel Superior (NS). Los cursos del NS y el NM se diferencian en cuanto a su alcance, pero ambos se evalúan según los mismos descriptores de calificaciones finales, con la salvedad de que se espera que los alumnos del NS demuestren un abanico más amplio de conocimientos, comprensión y habilidades.

Cada alumno debe cursar un mínimo de tres asignaturas (y un máximo de cuatro) del NS, y el resto del NM.

También existen cursos interdisciplinarios que cubren más de un grupo de asignaturas. Por ejemplo, Sistemas Ambientales y Sociedades permite satisfacer las exigencias del grupo Individuos y Sociedades y las del grupo Ciencias. Estos cursos interdisciplinarios aportan a los alumnos una mayor flexibilidad a la hora de elegir qué seis asignaturas quieren cursar.

ARTES

Modelo del Programa del Diploma: ilustración de la estructura del programa

PROGRAMA DEL DIPLOMA DEL BROUA

ESTUDIOS DE LENGUA

LITERATURA

ARTES

AR

Figura 64
Modelo del Programa del Diploma: ilustración de la estructura del programa

Existen reglas específicas para evitar que los alumnos cursen asignaturas con contenidos que se solapan. Dichas reglas se detallan en los *Procedimientos de evaluación* del PD.

## Modo de calcular el resultado del PD

Para calcular la puntuación total lograda en el PD, 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 de los componentes troncales. Así, la puntuación más alta posible que puede obtener un alumno es 45 puntos\*.

Este enfoque implica que las asignaturas de NM y NS tienen el mismo peso a la hora de determinar la puntuación final del alumno.

\* La puntuación máxima de 45 se obtiene de la siguiente manera: 6 (asignaturas) multiplicado por 7 (calificación máxima), más 3 (máximo de puntos de los componentes troncales).

#### Matriz de puntos para los componentes troncales

A diferencia de las demás asignaturas, Teoría del Conocimiento (TdC) y la Monografía reciben calificaciones alfabéticas, que van de la A a la E. El tercer componente troncal (Creatividad, Actividad y Servicio) no recibe una calificación, ya que no sería pertinente evaluar el desempeño en esta área.

Los componentes troncales aportan entre 0 y 3 puntos a la puntuación total lograda en el PD. Si el alumno obtiene una calificación de E en TdC o en la Monografía, o si no completa Creatividad, Actividad y Servicio, es posible que no obtenga el diploma. Para calcular la puntuación, se utiliza la tabla que aparece a continuación.

Figura 65

Matriz de puntos para los componentes troncales

	Teoría del Conocimiento (TdC)					
Monografía	Calificación final obtenida	А	В	С	D	E
	А	3	3	2	2	Condición excluyente
	В	3	2	2	1	
	С	2	2	1	0	
	D	2	1	0	0	
	E	Condición excluyente				

## **Condiciones excluyentes**

Un alumno no podrá recibir el diploma si se da alguna de las nueve condiciones siguientes:

- No se han cumplido los requisitos de CAS.
- La puntuación total no alcanza 24 puntos.
- Se ha obtenido una "N" (no se ha concedido ninguna calificación) en Teoría del Conocimiento, la Monografía o alguna de las asignaturas que cuentan para la obtención del diploma.
- Se ha obtenido la calificación "E" (elemental) en la Monografía, Teoría del Conocimiento o ambas.
- Se ha obtenido una calificación de 1 en cualquier asignatura y nivel.
- Se ha obtenido una calificación de 2 en tres ocasiones o más (NM o NS).
- Se ha obtenido una calificación de 3 o inferior en cuatro ocasiones o más (NM o NS).
- Se han obtenido menos de 12 puntos en las asignaturas de NS (para los alumnos matriculados en cuatro asignaturas de NS, contarán las tres calificaciones más altas).
- Se han obtenido menos de 9 puntos en las asignaturas de NM (los alumnos matriculados en dos asignaturas de NM deberán obtener al menos 5 puntos en este nivel).

### Diploma bilingüe

En lugar del diploma del IB habitual, el alumno puede obtener un diploma bilingüe si cumple al menos uno de los siguientes criterios:

- Haber completado dos lenguas del Grupo 1 con un mínimo de una calificación final de 3 en ambas.
- Haber completado una asignatura del Grupo 3 o del Grupo 4 en una lengua distinta de su lengua del Grupo 1. El alumno debe obtener una calificación final de 3 o superior tanto en la lengua del Grupo 1 como en la asignatura del Grupo 3 o del Grupo 4.

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
- Un programa de estudios del colegio
- Una asignatura adicional a las seis asignaturas conducentes al diploma

# Método para determinar las puntuaciones ausentes

En algunos casos no es posible obtener o utilizar el trabajo de un alumno. Esto sucede, por ejemplo, si un trabajo se pierde o se destruye durante el envío por correo postal, o si hay pruebas de la existencia de conducta improcedente a gran escala pero no es posible identificar a qué trabajos ha afectado. En tales circunstancias, el IB puede intentar calcular la calificación que se habría obtenido, para lo cual se utiliza el procedimiento de ausencia de puntuación.

El proceso general, con sus puntos fuertes y débiles, se describe en la sección "Casos de ausencia de puntuación". A continuación se describe el cálculo que se utiliza en el PD.

Si en un colegio hay cinco alumnos o más que cursen la asignatura, se utilizan los datos del colegio para determinar las puntuaciones ausentes. Si en un colegio hay menos de cinco alumnos que cursen la asignatura, se utilizan datos mundiales para calcular las puntuaciones ausentes. La experiencia apunta a que, si el colegio tiene cinco alumnos o más, suele haber pocas diferencias en los resultados entre utilizar los datos del colegio o los datos mundiales.

#### Figura 66

Cómo determinar la puntuación ausente de un alumno

Para determinar la puntuación ausente de un alumno:

Puntuación total ajustada del alumno en los otros componentes

Cociente =

Puntuación total ajustada mundial (o del colegio) en los otros componentes

Puntuación ausente del alumno = Puntuación ajustada mundial (o del colegio) en el componente que falta x Cociente

El cociente compara las puntuaciones del alumno con la media mundial o del colegio en los componentes que dicho alumno ha completado:

- Si el alumno ha obtenido mejores resultados que la media, esta cifra será mayor que 1.
- Si el alumno 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 de un alumno y, si es necesario, aplicar los procesos de consideración especial.

# Gestión de los estándares en grupos de alumnos de distintos tamaños

Las asignaturas del Programa del Diploma pueden tener una cantidad de alumnos muy dispar, desde solo 1 hasta más de 40.000. Todas las asignaturas son igual de importantes y el uso de las calificaciones será el mismo en los grupos con solo un alumno que en los grupos con miles. Sin embargo, por razones prácticas, se utilizan procesos ligeramente distintos en las asignaturas en función del número de alumnos matriculados.

#### Mantenimiento de la calidad del trabajo de los examinadores

El modelo de calidad (que implica utilizar exámenes de práctica, de aptitud y de control) está pensado para garantizar que los examinadores entiendan y apliquen el estándar del examinador principal. Sin embargo, si hay muy pocos examinadores, puede que no resulte adecuado emplear este modelo.

Si todos los examinadores participan en la discusión que se realiza en el momento de la estandarización, no hará falta que haya exámenes de práctica ni de aptitud, puesto que todos los examinadores ya habrán participado en el proceso de establecer del estándar. No obstante, se seguirán utilizando exámenes de control para garantizar que los examinadores continúan corrigiendo según el estándar previsto.

Como siempre existe el riesgo de que tener que introducir más examinadores si la corrección está siendo lenta, normalmente se prepara un modelo de calidad completo, aunque finalmente no haga falta utilizarlo. En el caso de que no se hayan preparado exámenes de aptitud, se pueden utilizar exámenes de control en su lugar, si bien no es lo ideal, ya que se tienen en consideración elementos distintos al seleccionar estos dos tipos de exámenes corregidos definitivos.

Lo más importante es mantener la calidad de la corrección, aunque una asignatura tenga muy pocos alumnos. En ese sentido, el IB utiliza todas las pruebas disponibles para detectar cualquier problema antes de la publicación de los resultados.

## Calificaciones: proceso completo de concesión de calificaciones (virtual o presencial)

En el caso de las asignaturas en las que hay varios examinadores por debajo del examinador principal, se realiza una reunión formal para aplicar las prácticas que se describen en la sección "Concesión de calificaciones finales (y totalización)". En esta reunión habrá un responsable de la asignatura del IB para ayudar al examinador jefe y realizar un control de calidad del proceso.

Estas reuniones pueden ser presenciales o virtuales. Si son virtuales, es importante tener en cuenta que los examinadores supervisores pueden estar en distintas zonas horarias. En esas reuniones, los examinadores supervisores utilizan un foro en línea privado para compartir sus puntos de vista mientras estudian las pruebas disponibles; además, realizan videoconferencias para consensuar las decisiones clave.

## Calificaciones: proceso guiado de concesión de calificaciones en asignaturas con pocos alumnos (virtual)

En los casos en los que hay muy pocos examinadores en cada componente, los examinadores supervisores se reúnen virtualmente para discutir. El 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 progreso del trabajo para garantizar la calidad, si bien es poco probable que participe en las discusiones. Los equipos de examinadores supervisores encargados de la concesión de calificaciones en asignaturas con pocos alumnos también suelen tener pocos integrantes.

### Calificaciones: proceso clásico de concesión de calificaciones en asignaturas con pocos alumnos

Cuando el número de alumnos es demasiado bajo como para que las pruebas estadísticas acumuladas con los años sean significativas, se pide a los examinadores que revisen todos los trabajos de los alumnos en vez de revisar solo una muestra para recomendar los límites de calificación estimativos. En esos casos, los examinadores principales y los examinadores jefe pueden recomendar todos los límites de calificación, y no solo los límites 2-3, 3-4 y 6-7. Esto se debe a que con frecuencia es posible que no haya trabajos de alumnos en un límite de calificación concreto. Esta situación se ilustra en el siguiente ejemplo:

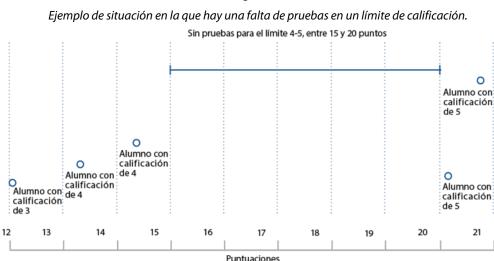


Figure 67

Los 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. Suele ser difícil mantener un estándar coherente en estas asignaturas con pocos alumnos, ya que la poca cantidad de alumnos limita el valor de los datos históricos. Además, con frecuencia estos alumnos se encuentran en la parte superior del rango de puntuaciones.

Para facilitar el trabajo de los examinadores de todas las asignaturas con pocos alumnos (ya sea en el proceso guiado o en el proceso clásico de concesión de calificaciones), el IB recomienda que se realicen discusiones entre los examinadores 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 cuántos alumnos hayan realizado los exámenes, la última etapa del proceso de concesión de calificaciones es que el examinador jefe haga recomendaciones al IB y que el personal de evaluación sénior confirme que las pruebas que respaldan esas recomendaciones les convencen.

El IB tiene en cuenta la importancia relativa de las pruebas estadísticas en función del número de alumnos del grupo, pero todas las asignaturas se someten al mismo escrutinio.

#### Lecturas adicionales

Para obtener más información sobre el Programa del Diploma del IB, puede consultar los siguientes recursos:



- El Programa del Diploma: de los principios a la práctica
- Reglamento general del Programa del Diploma
- Procedimientos de evaluación del Programa del Diploma
- Normas para los Colegios del Mundo del IB: Programa del Diploma
- 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)

# El Programa de Orientación Profesional del IB

Las características distintivas de la evaluación del POP son:

- Los alumnos deben cumplir una serie de requisitos para conseguir el certificado del POP.
- No existe una puntuación total para el certificado del POP.
- Cuando los alumnos del POP estudian cursos junto con alumnos del PD, se realiza una evaluación común.
- El marco del POP exige que los alumnos realicen unos estudios de formación profesional que el IB no ofrece y no evalúa.

# Objetivos generales del Programa de Orientación Profesional

La validez de los resultados de la evaluación solo se puede determinar si se tiene claro el propósito del curso y del programa. Por tanto, esta sección comienza con una presentación de los objetivos generales del programa.

La particularidad del POP es que prepara a los alumnos para ingresar en el mundo laboral, sea cual sea la carrera profesional que elijan. El programa asegura que los alumnos desarrollen habilidades transferibles que les servirán para toda la vida en el empleo para el que se quieran capacitar.

El programa permite a los alumnos:

- Desarrollar una variedad de competencias laborales generales y profundizar su comprensión en áreas de conocimiento específicas mediante los cursos del Programa del Diploma que estudien
- Desarrollar estrategias flexibles para la adquisición y profundización de conocimientos en diversos contextos
- Prepararse para una participación eficaz en el cambiante mundo laboral
- Fomentar aptitudes y hábitos mentales que les permitan mantener una actitud de aprendizaje durante toda la vida, con la voluntad de considerar nuevas perspectivas
- · Participar en un aprendizaje que desarrolle su capacidad y voluntad de generar un cambio positivo

El Programa de Orientación Profesional: de los principios a la práctica, 2015.

#### 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 del POP y el certificado se puedan utilizar para determinar los siguientes aspectos:

- · La selección para oportunidades laborales o programas de empleo, como prácticas profesionales
- La selección para estudios posteriores en el área de formación profesional correspondiente
- La selección para la entrada en la universidad
- Si los alumnos ya han cumplido algunos requisitos de algún programa universitario (si se les pueden otorgar créditos o validar alguna asignatura o curso)

En los casos en los que el alumno ha realizado la evaluación en una lengua (de respuesta) concreta, eso también demuestra que puede cursar estudios posteriores de esa materia o profesión en dicha lengua.

Para obtener más información sobre la importancia de estos usos válidos, consulte la sección "¿Qué se considera válido: la evaluación o su uso?".



## La estructura del POP

El POP es un marco educativo compuesto por tres partes, que son:

- Al menos dos cursos del PD de Nivel Medio (NM) o Nivel Superior (NS)
- Los componentes troncales del POP
- Estudios de formación profesional

Figura 68 Modelo del Programa de Orientación Profesional: ilustración de la estructura del programa



## Los componentes troncales, incluido el Proyecto de Reflexión

Los componentes troncales están pensados para contextualizar tanto los cursos del PD como los estudios de formación profesional. Los componentes troncales son el nexo de unión entre todas las áreas de aprendizaje.

Es obligatorio completar los cuatro componentes troncales, que son:

- Habilidades Personales y Profesionales
- Aprendizaje-servicio
- Desarrollo de la Lengua
- Proyecto de Reflexión

El IB solamente evalúa el Proyecto de Reflexión, que corrige un profesor y luego modera el IB. Con respecto a los demás elementos, el colegio debe confirmar al IB que el alumno los ha completado satisfactoriamente, pero el IB no comprueba la evaluación (si la hay).

#### Cursos del Programa del Diploma como parte del POP

Todos los alumnos tienen que cursar como mínimo dos asignaturas del PD. A los alumnos del POP y a los del PD se les aplica el mismo proceso de evaluación. Es decir, ningún examen y ningún proceso de concesión de calificaciones finales es específico del POP.

A pesar de que hay asignaturas que son comunes para el PD y el POP, no está permitido que un alumno curse 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 ni evalúa ni controla la calidad de los resultados de la parte de estudios de formación profesional del POP. El único requisito es que el colegio confirme que el alumno 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 cumple todos los requisitos que se detallan a continuación, se le concederá el certificado del POP.

- El colegio ha confirmado que el alumno ha completado los estudios de formación profesional específicos.
- Haber obtenido, como mínimo, un 3 en al menos dos de los cursos del Programa del Diploma.
- Haber obtenido, como mínimo, una D en el Proyecto de Reflexión.
- El colegio ha confirmado que se han cumplido todos los requisitos de Habilidades Personales y Profesionales, Aprendizaje-servicio y Desarrollo de la Lengua.
- No haber recibido una sanción del Comité de la evaluación final por conducta improcedente.

Las calificaciones obtenidas en las asignaturas del PD y en el Proyecto de Reflexión las confirma el mismo Comité de la evaluación final que para el PD.

## Certificado bilingüe del POP

En lugar del certificado del POP habitual, el alumno puede obtener un certificado bilingüe 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 debe haber obtenido como mínimo una calificación de 3 en ambos cursos.

# Gestión de los estándares y trabajos faltantes

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 alumnos.

## Lecturas adicionales

Para obtener más información sobre el Programa de Orientación Profesional del IB, 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 Desarrollo de la Lengua
- Guía de Habilidades Personales y Profesionales
- Guía de Aprendizaje-servicio

# El Programa de los Años Intermedios del IB

Las características distintivas de la evaluación del PAI son:

- Todos los alumnos en el quinto año deben completar el Proyecto Personal, que se modera externamente, pero las demás evaluaciones del IB son opcionales para los colegios que imparten el PAI.
- Los alumnos deben cursar una cantidad obligatoria de asignaturas para obtener el certificado del PAI.
- El logro total alcanzado en el certificado del PAI se describe mediante una puntuación cuyo máximo es de 56 puntos.
- Los requisitos troncales de aprendizaje interdisciplinario y Proyecto Personal contribuyen a la puntuación total lo mismo que las otras asignaturas. El servicio comunitario no aporta ninguna puntuación que contribuya al total.
- En lo que respecta a las evaluaciones sumativas diseñadas por el IB:
- Cada asignatura tiene solo un componente
- Las asignaturas se evalúan mediante una carpeta electrónica o exámenes en pantalla

# Objetivos generales del Programa de los Años Intermedios

La validez de los resultados de la evaluación solo se puede determinar si se tiene claro el propósito del curso y del programa. Por tanto, esta sección comienza con una presentación de los objetivos generales 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.

En el PAI, los objetivos específicos de los grupos de asignaturas se corresponden claramente con los criterios de evaluación. 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 la enseñanza y el aprendizaje 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.



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 de cada examen en pantalla de carácter disciplinario estarán relacionadas con el contexto global elegido, o inspiradas o basadas en él. La totalidad del examen en pantalla acerca del aprendizaje interdisciplinario se basará en el contexto global elegido.

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 de los alumnos que continúan con sus estudios

En los casos en los que el alumno 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 español, francés o inglés).

Para obtener más información sobre la importancia de estos usos válidos, consulte la sección "¿Qué se considera válido: la evaluación o su uso?".

## La estructura del PAI

Figura 69 Modelo del Programa de los Años Intermedios: ilustración de la estructura del programa



# Modo de calcular el resultado del PAI

Al final del quinto año del PAI, los alumnos pueden matricularse en la evaluación externa del IB. Los resultados de estas evaluaciones quedarán registrados en el documento de resultados de los cursos del PAI. Además, estos alumnos pueden decidir si quieren realizar las evaluaciones con las que se les podría conceder el certificado del PAI.

El colegio también puede emitir un título de participación del PAI, pensado para los alumnos del PAI que cursan el programa durante dos años como mínimo y cumplen con los requisitos correspondientes al tercer o cuarto año. Estos alumnos no están matriculados para ningún tipo de evaluación con el IB. El título de participación es un documento del colegio y no está verificado por el IB.

Para conseguir el certificado del PAI, el alumno debe participar en 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 dichas asignaturas
- Completar el examen en pantalla de la evaluación interdisciplinaria y obtener una calificación de 3, como mínimo



- Completar el Proyecto Personal con una calificación de 3, como mínimo
- Obtener una puntuación total de 28 puntos, como mínimo
- Cumplir las expectativas de servicio comunitario del colegio

El certificado bilingüe del PAI requiere además obtener resultados satisfactorios en las evaluaciones 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 que no sea la que el alumno eligiera para el curso de Lengua y Literatura

# Desarrollo de la evaluación sumativa externa (evaluación electrónica del PAI)

La evaluación electrónica opcional contempla dos formas de determinar los conocimientos y las competencias de los alumnos:

- Carpetas electrónicas que contienen trabajos de alumnos de Adquisición de Lenguas, Artes, Diseño, y Educación Física y para la Salud, que posteriormente se someten a moderación para garantizar que se sique un estándar global coherente
- Exámenes en pantalla (dos horas de duración) para los cursos de Lengua y Literatura, Individuos y Sociedades, Ciencias, y Matemáticas, así como para el aprendizaje interdisciplinario

Además, el Proyecto Personal se envía electrónicamente al IB para su moderación. Si bien hay otros componentes de la evaluación electrónica que son opcionales para los colegios, todos los alumnos del quinto año del PAI deben 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 alumnos comprender la naturaleza y el propósito de la evaluación electrónica del PAI. También ayudan a los alumnos a prepararse para los exámenes en pantalla y a centrarse en los criterios y las estrategias de evaluación de cada grupo de asignaturas. 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 tres puntos.

## Carpetas electrónicas y planificadores de unidades parcialmente completados

Las carpetas electrónicas permiten evaluar el desempeño o las tareas (productos) de trabajos de clase realizados durante un largo período de tiempo, lo cual, debido a su naturaleza, resultaría muy difícil de evaluar mediante un examen. 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 a los profesores a asegurarse de que los trabajos de los alumnos 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 encarguen los profesores permitan a los alumnos alcanzar todas las calificaciones del PAI. Las evaluaciones que diseñan los profesores implican el riesgo de poner a los alumnos 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 disponible del alumno; si la tarea que diseña el profesor 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 las cargas de trabajo para los alumnos

Para los alumnos, 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 que, en conjunto, es más adecuado minimizar la cantidad de evaluación sumativa para los alumnos de 16 años que cursan el PAI. Si bien esto provoca dificultades porque los alumnos solo tienen una oportunidad para demostrar sus logros, el IB las acepta para procurar el bienestar general de los alumnos.

# Método para determinar las calificaciones ausentes

Uno de los inconvenientes de querer reducir la carga de trabajo de los alumnos 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 es menos fiable que en los otros programas, por lo cual solo se debe utilizar en casos excepcionales.

- Si un alumno no ha realizado una evaluación, se le debe ofrecer la oportunidad de realizar el examen en una convocatoria posterior.
- Si un alumno ha realizado una evaluación pero ha habido factores externos que probablemente hayan perjudicado su desempeño, se deben aplicar consideraciones especiales.
- Si un alumno 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 o del colegio, se aplica el procedimiento de ausencia de calificación.

Para que se pueda aplicar el procedimiento de ausencia de puntuación, el alumno debe haber recibido una calificación final en al menos otros cuatro cursos del PAI. Es imposible calcular resultados para los alumnos que tengan 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 del alumno 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 de un alumno y, si es necesario, aplicar los procesos de consideración especial.

## Lecturas adicionales

Para obtener más información sobre el Programa de los Años Intermedios del IB, 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
- Guía de los proyectos del PAI
- Procedimientos de evaluación del Programa de los Años Intermedios
- Requisitos de TI para llevar a cabo los exámenes en pantalla del PAI
- Familiarización con los exámenes en pantalla del PAI destinada a los alumnos (PC y Mac)



# El Programa de la Escuela Primaria del IB

Las características distintivas de la evaluación del PEP son:

- La evaluación implica la colaboración de maestros y alumnos para el seguimiento, la documentación, la medición, la comunicación de resultados y la adaptación del aprendizaje.
- El IB no impone ninguna evaluación sumativa externa.

# Objetivos generales del Programa de la Escuela Primaria

La validez de los resultados de evaluación solo se puede determinar si se tiene claro el propósito del programa. Por tanto, esta sección comienza con una presentación de los objetivos generales del programa.

El PEP se centra en el corazón y en la mente de los alumnos, y atiende sus necesidades sociales, físicas, emocionales y culturales además de aquellas consideradas más académicas. Se valoran las áreas disciplinarias tradicionales al tiempo que se hace hincapié en el equilibrio entre la adquisición de conocimientos y habilidades esenciales, y la búsqueda de significados y comprensión en relación con el mundo que los rodea. El programa ofrece a los alumnos la oportunidad de construir significados, principalmente a través de la indagación impulsada por conceptos. Los distintos componentes del aprendizaje se integran en el programa de indagación transdisciplinario, que a su vez les permite establecer relaciones con la vida fuera del colegio.

Documentación del PEP del IB

## La estructura del PEP

Figura 70

Modelo del Programa de la Escuela Primaria: ilustración de la estructura del programa



# La evaluación formativa en apoyo del aprendizaje

En el PEP, el objetivo principal de la evaluación es proporcionar información acerca de la marcha del proceso de aprendizaje. Bruner sostiene que la evaluación no debe representar un modo de recompensar o castigar al alumno, sino una vía para ofrecerle información (Bruner, 1961). Los maestros deben seleccionar estrategias y diseñar instrumentos de evaluación que reflejen claramente los objetivos de aprendizaje concretos sobre los que desean informar, y emplear una variedad de técnicas que tomen en cuenta las diversas, complejas y sofisticadas maneras en que los alumnos logran comprender sus experiencias. Asimismo, el PEP hace hincapié en la autoevaluación y reflexión por parte del alumno y del maestro.

## Lecturas adicionales

Para obtener más información sobre el Programa de la Escuela Primaria del IB, puede consultar la página de recursos del PEP en el Centro de recursos para los programas.



## Anexo 1: Moderación de la evaluación interna

La moderación se utiliza en los trabajos de evaluación interna para garantizar que haya 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 los criterios de evaluación al corregir el trabajo de los alumnos.

### Muestras

- Los profesores de un colegio deben seguir un estándar común en la corrección, ya que, si es necesario, se aplica un factor de moderación para todos los alumnos de una asignatura.
- Para la muestra para la moderación, el IB no suele seleccionar a alumnos que hayan logrado la máxima puntuación, ya que así permite que los alumnos con el rango de puntuaciones más elevadas tengan la posibilidad de que sus trabajos se moderen al alza.

La muestra de la evaluación interna se selecciona cuidadosamente para garantizar que el rango de puntuación 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 matriculados en la asignatura. El IB selecciona a los alumnos cuyos trabajos conforman las muestras para la moderación de la evaluación interna después de que el colegio haya entregado formalmente las puntuaciones de la evaluación interna. Para la muestra para la moderación, el IB no suele seleccionar a alumnos que hayan logrado la máxima puntuación, ya que así permite que los alumnos con el rango de puntuaciones más elevadas tengan la posibilidad de que sus trabajos se moderen al alza.

Cuando en un colegio hay suficientes alumnos matriculados en un curso determinado como para que puedan repartirse en más de una clase, y más de un profesor realiza la evaluación interna, el IB espera que los profesores 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 de alumnos corregidos por los distintos profesores. 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 tiene acceso a más cantidad de información que el examinador sobre el contexto y el proceso que rodea al trabajo de un alumno. Por ese motivo, a los examinadores que moderan los componentes de evaluación interna se les pide que juzguen si la corrección del profesor resulta adecuada; no se trata de volver a corregir el trabajo y rechazar las puntuaciones del profesor. De hecho, esas puntuaciones del profesor solamente se pueden modificar si el moderador está seguro de que resultan inadecuadas.

El examinador modera la muestra de trabajos y hace un ajuste a las puntuaciones del profesor para todos los alumnos del colegio en ese componente; para ello se basa en una comparación estadística entre los dos grupos de puntuaciones (utilizando una regresión lineal).

- Si el profesor puntúa por exceso o por defecto de manera repetida, se realizará el mismo ajuste en todas las puntuaciones que otorgue el profesor.
- Si el profesor puntúa por exceso o por defecto 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.

Si el profesor está empleando el estándar correcto en su corrección, no se realizará ningún ajuste.

## Regresión lineal

En cada muestra para la moderación se analizan los datos, lo cual permite aplicar un ajuste adecuado a todas las puntuaciones de un profesor 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 y el examinador han concedido a los trabajos de la muestra. A continuación se muestra un ejemplo de regresión lineal y su correspondiente conjunto de puntuaciones.

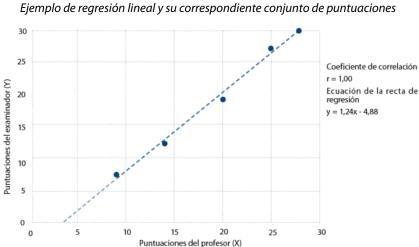


Figura AN1
Ejemplo de regresión lineal y su correspondiente conjunto de puntuaciones

Recta de regresión de la moderación para un profesor que es levemente demasiado severo en el extremo superior y demasiado generoso en el inferior. Cada uno de los puntos representa las dos puntuaciones concedidas a un trabajo de muestra: una del profesor y la otra del examinador. La recta de regresión continua se utiliza para convertir las puntuaciones del profesor en puntuaciones moderadas.

Figura AN2

Las puntuaciones correspondientes que conceden el profesor y el examinador se comparan mediante una regresión lineal. La recta calculada que mejor se ajusta a la situación, o "factor de moderación", da las puntuaciones moderadas finales.

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 en una puntuación (y) equivalente que el examinador probablemente hubiera concedido a ese mismo alumno. 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. Las variaciones relacionadas con alumnos específicos no se pueden tomar en consideración. El propósito de la moderación es garantizar que, en

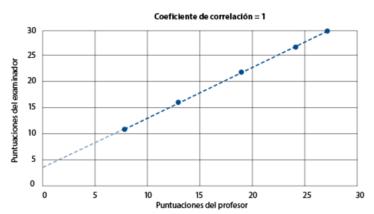
conjunto, las puntuaciones de los alumnos se ajustan a niveles más adecuados. La moderación no puede garantizar un resultado exactamente correcto para cada alumno.

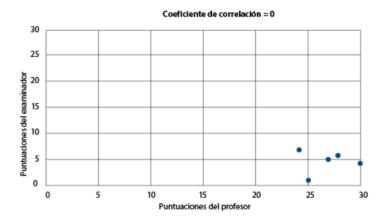
### Error en la moderación

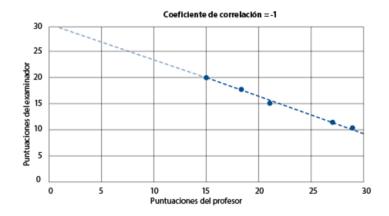
Automáticamente se realiza una comprobación para asegurarse de que la recta de regresión lineal (es decir, el factor de moderación calculado) cumple determinadas condiciones antes de aplicarse a todas las puntuaciones de un profesor. En algunos casos, tal vez no sea posible calcular un ajuste de moderación 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 y las del examinador.

- Un coeficiente de correlación de 0 indica que no hay ninguna relación.
- Un coeficiente de 1 indica que hay una coherencia perfecta en la relación entre las puntuaciones, y que hay coincidencia en la clasificación de los alumnos de mejor a peor (aunque no necesariamente las mismas puntuaciones).
- Un coeficiente de -1 indica que hay un desacuerdo constante en relación con los méritos relativos del trabajo de los alumnos, y que el profesor y el examinador producen clasificaciones opuestas.

Figura AN3
Gráficos del coeficiente de correlación







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 acuerdo entre el profesor y el examinador. 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 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 ha tenido que comprimir considerablemente el rango de puntuaciones del

profesor. Si el gradiente es superior a 1,5, la recta es demasiado inclinada y sucede lo contrario: el profesor no ha diferenciado lo suficiente entre el trabajo de menor y de mayor calidad de los alumnos, y el examinador 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, tienen en cuenta los datos subyacentes y pueden decidir que:

- La recta de regresión calculada es adecuada para el rango de puntuaciones concreto del profesor.
- Hay que aplicar otro ajuste de moderación que resulte adecuado para el rango de puntuaciones del profesor.
- Hay que solicitar más datos de muestras para poder aclarar la tendencia. 3.
- Hay que solicitar los trabajos de los alumnos restantes para poder realizar una revisión completa de la corrección del profesor.

Todos los trabajos de evaluación interna de los alumnos deben estar disponibles hasta que se publiquen los resultados, para que se 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". Se observa que el ajuste obtenido mediante una línea recta puede producir efectos inapropiados en los extremos de la gama de puntuaciones posibles, de manera que resulta imposible que un alumno obtenga un 0. Es posible que a un alumno 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 cuyo trabajo es de poca calidad, pero aun así merezca algunos puntos, reciba un 0 después de la moderación, como se refleja en la figura AN3.

Para superar este problema, se aplica 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, como se ve en la figura AN4. El truncamiento de la recta de regresión no se aplica en el extremo superior, ya que a menudo los profesores conceden la puntuación máxima a alumnos que claramente no la merecen.



Figura AN4
Ejemplo de truncamiento en una regresión lineal

"Truncamiento" de una recta de regresión para evitar que las puntuaciones de los alumnos se ajusten al alejarse mucho de los valores mínimos o al acercarse mucho a estos.

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 supera el proceso de moderación automática. Si la puntuación del profesor no supera la moderación y no se puede moderar automáticamente, no se aplica el truncamiento.

# Anexo 2: Hoja de ruta para crear un argumento de validez

El IB debe asegurarse de que sus cursos y evaluaciones sean adecuados al propósito. Para esto no basta con completar una lista de verificación, sino que se requiere una discusión concienzuda acerca de cómo hallar un equilibrio entre las distintas prioridades contrapuestas y qué pruebas demuestran que se han cumplido los objetivos y las intenciones.

El argumento de validez conforma el resumen de todas estas discusiones, ya que reúne los distintos aspectos que explican por qué el IB considera que sus cursos y evaluaciones son lo mejor posible. Los temas y las preguntas que aparecen a continuación forman el punto de partida para demostrar la validez. Algunas respuestas tienen que ver con procesos generales en la evaluación del IB, como los exámenes de aptitud y de control, pero otras de las respuestas se refieren a asignaturas y cursos específicos.

El proceso de creación de un argumento de validez no es un fin en sí mismo: lo importante es hacer las preguntas correctas y registrar pruebas en cada una de las partes del ciclo para que las evaluaciones y los cursos del IB sean de la mayor calidad posible.

#### Desarrollo del curso correcto

- ¿El currículo resulta lo suficientemente adecuado y contemporáneo como para que los alumnos puedan entender el curso?
- ¿En qué se diferencia el curso de otros cursos del IB?
- ¿Cómo sabe el IB que el currículo se puede impartir de manera eficaz en el tiempo disponible para la enseñanza?
- En el proceso de revisión del currículo del IB se recaban pruebas para responder estas preguntas relativas al argumento de validez.

#### Alineación de los objetivos de la evaluación con los objetivos generales del curso

¿Qué significa ser bueno en esta asignatura? ¿Cómo se demuestra que se es bueno?

¿Qué se intenta evaluar con cada tarea (componente)?

¿En qué medida coinciden los objetivos generales del curso con lo que se evalúa?

¿De qué forma los criterios de evaluación y los descriptores de calificaciones finales reflejan cómo es una buena respuesta para cada tarea?

Entre los ejemplos de pruebas cabe destacar las explicaciones y justificaciones del modo de evaluación y las notas sobre discusiones acerca de los criterios y las descripciones.

#### Valoración del modelo de evaluación

¿Cómo sabe el IB que los estándares curriculares y los de evaluación (no los estándares de desempeño) son equiparables con otros cursos del IB?

¿Cómo minimiza los sesgos el modelo de evaluación?

¿Cómo se pueden adaptar las evaluaciones para hacerlas aptas para alumnos con requisitos especiales (por ejemplo, los alumnos ciegos o los que no pueden realizar una tarea concreta)?

¿Qué nivel de desacuerdo profesional resulta aceptable al corregir?

Es posible encontrar pruebas durante la revisión del currículo y en los informes generales de la asignatura.

#### Reflexión sobre el proceso de creación de los exámenes (exámenes de muestra y exámenes reales)

¿De qué manera las preguntas reflejan con precisión el currículo?

Entre los ejemplos de pruebas cabe destacar la tabla de elaboración (que describe cómo están relacionadas las preguntas con el currículo y con los objetivos de evaluación), y los informes de los verificadores.

#### **Disposiciones especiales**



¿Las adecuaciones de la evaluación hicieron que fuera accesible a una amplia variedad de alumnos sin que estos debieran solicitar disposiciones especiales?

En los casos en que fueron necesarias disposiciones especiales, ¿hubo algún problema que no se hubiera identificado durante la fase de diseño?

¿Cómo se sabe si las disposiciones especiales aplicadas permitieron mantener la equiparabilidad del estándar?

Entre los ejemplos de pruebas cabe destacar las decisiones que condujeran a la creación de exámenes modificados o la evaluación del desempeño basada en preguntas modificadas.

#### Manejabilidad al realizar la evaluación

¿Hubo algún problema con los colegios o con los alumnos que fuera resultado de los procesos del IB o del diseño de la evaluación?

Entre los ejemplos de pruebas cabe destacar comentarios de los colegios o los alumnos. La ausencia de comentarios puede indicar que no hay problemas.

#### Revisión de una corrección satisfactoria

¿En qué medida fue fiable la corrección? ¿Los márgenes de tolerancia podrían haber sido más reducidos?

¿El esquema de calificación inicial (realizado con el examen) resultó en su mayoría correcto, o hubo que hacer modificaciones importantes? ¿Las modificaciones fueron razonables, o durante la fase de creación de los exámenes (y de los esquemas de calificación) se podrían haber previsto las respuestas de los alumnos?

Entre los ejemplos de pruebas cabe destacar los datos de los exámenes de aptitud y de control, y cualquier sustitución de exámenes corregidos definitivos.

#### Confianza y equiparabilidad en las decisiones de concesión de calificaciones

¿En qué medida se sentía seguro el examinador jefe a la hora de recomendar los límites de calificación?

¿Todas las pruebas respaldaban la decisión respecto a los límites de calificación, o distintos tipos de información apuntaban a conclusiones distintas?

¿Cómo sabe el IB si las calificaciones resultaron coherentes con las de años anteriores y con las de otras asignaturas?

Entre los ejemplos de pruebas cabe destacar los comentarios contenidos en los informes internos.

#### Pruebas de equidad (sesgos)

¿Cómo sabe el IB si se trató por igual a los distintos grupos de alumnos?

¿En qué medida el IB está seguro de que la corrección y la calificación fueran coherentes entre las diferentes lenguas de respuesta y zonas horarias?

Entre los ejemplos de pruebas cabe destacar los análisis de los resultados y los comentarios de los profesores. También se incluirían pruebas que aporten los examinadores que corrigen en español y en francés y que participan en la concesión de calificaciones finales, datos sobre los examinadores bilingües que hayan corregido exámenes de control en inglés, y la revisión del trabajo de los alumnos de otras zonas horarias al establecer límites de calificación.

#### Consecuencias de las consultas sobre los resultados

¿En qué medida el IB está seguro de que los cambios en la puntuación a raíz de las consultas sobre los resultados son razonables?

¿En qué medida el IB está seguro de que las consultas sobre los resultados demuestran que no hubo ningún error sistemático en la corrección o la calificación?

Entre los ejemplos de pruebas cabe destacar el análisis de cualquier patrón de cambios en la consulta sobre los resultados, y pruebas de que la cantidad de pequeños cambios de puntuación al alza y a la baja es la misma. El proceso del IB de utilizar exámenes de control en las consultas sobre los resultados también sirve de prueba. También resultan pertinentes las pruebas de la implicación de los examinadores supervisores en este proceso para asegurarse de que se respeta el estándar.

#### Validez predictiva



¿Qué pruebas hay de que los resultados de los alumnos sirven como un buen indicador del éxito futuro en esa asignatura? ¿Qué pruebas hay de que sea un buen indicador del éxito en las asignaturas relacionadas? Entre los ejemplos de pruebas cabe destacar los comentarios cualitativos de los alumnos y del personal, así como las pruebas cuantitativas que obtenga el equipo de investigación del IB.

#### Pruebas de una cultura de mejora continua

¿Qué aprendió el IB de la convocatoria anterior para mejorar la actual?

¿Qué lecciones clave puede extraer el IB de la convocatoria actual que sirvan para futuras convocatorias?

¿Cómo evalúa el IB el éxito en la implementación de estas mejoras?

Entre los ejemplos de pruebas cabe destacar la manera en que las observaciones de convocatorias anteriores influyeron en el diseño de la actual.

## Bibliografía

Anderson, L. W. y Krathwohl, D. R. A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Nueva York (EE. UU.): Longman, 2001.

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.

Black, P. "Assessment, Learning Theories and Testing Systems". En Murphy, P. Learners, Learning & Assessment. Londres (Reino Unido): Paul Chapman Publishing en asociación con The Open University. 1999.

Bloom, B. S.; Englehart, M. D.; Furst, E. J.; Hill, W. H. y Krathwohl, D. R. *Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain.* Nueva York (EE. UU.): Longman, 1956.

Broadfoot, P. Education, Assessment and Society: A Sociological Analysis. Buckingham (Reino Unido): Oxford University Press, 1994.

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 national and international schools.* Londres (Reino Unido): Kogan Page, 2002.

Chamberlain, S. "AQA – Public Perceptions of Reliability". En Ofqual. *Ofqual's Reliability Compendium*. Capítulo 18. Capítulo 18. Coventry (Reino Unido): 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]). Durham (Reino Unido): CEM Centre, Durham University, 2008.

Cresswell, M. J. "Examination Grade: How many should there be?". En *British Educational Research Journal*. 1986, vol. 12, n.º 1.

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*. Chichester (Reino Unido) y Nueva York (EE. UU.): Wiley, 1996. Pp. 57-84.

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.

Dolan, R. P.; Burling, K.; Harms, M.; Strain-Seymour, E.; Way, W. y Rose, D. H. A Universal Design for Learning-based framework for designing accessible technology-enhanced assessments. (Informe de investigación). lowa City, Iowa (EE. UU.): Pearson Education Measurement, 2013.

Frith, D. S. y Macintosh, H. G. *A Teachers' Guide to Assessment*. Cheltenham (Reino Unido): Stanley Thomas, 1984.

Gibbs, G. Assessing More Students. Oxford (Reino Unido): Oxford Centre for Staff Learning and Development, 1992

Gipps, C. y Murphy, P. A Fair Test? Assessment, Achievement and Equity. Buckingham (Reino Unido): Oxford University Press, 1994.

Glaser, R. "Instructional technology and the measurement of learning outcomes: Some questions". En *American Psychologist.* 1963, vol. 18. Pp. 519-521.

Goldstein, H. "Group differences and bias in assessment". En Goldstein, H. y Lewis, T. (eds.). *Assessment: Problems, Developments and Statistical Issues*. Chichester (Reino Unido) y Nueva York (EE. UU.): Wiley, 1996. Pp. 85-93.



Good, F. J. y Cresswell, M. J. Grading the GCSE. Londres (Reino Unido): 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. Coventry (Reino Unido): Office of Qualifications and Examinations Regulation, 2010. Pp. 68.

Hieronymous, A. N. y Hoover, H. D. lowa tests of basic skills manual for school administrators. lowa City (EE. UU.): University of Iowa, 1986.

Hughes, D.; Keeling, B. y Tuck, B. "Effects of achievement expectations and handwriting quality on scoring essays". En Journal of Educational Measurement. 1983, vol. 20, n.º 1. Pp. 65-70.

Humphreys, L. G. "An analysis and evaluation of test and item bias in the prediction context". En Journal of Applied Psychology. 1986, vol. 71. Pp. 327-333.

Lambert, D. y Lines, D. Understanding Assessment. Londres (Reino Unido): Routledge Falmer, 2000.

Llewellyn, D. et al. Inquire within: Implementing inquiry and argument-based science standards in grades 3–8. 3.ª ed. Thousand Oaks (EE. UU.): Corwin, 2013.

Linn, M. C. "Gender differences in educational achievement". En Pfleiderer, J. (ed.). Sex Equity in Educational Opportunity, Achievement, and Testing. Princeton, Nueva Jersey (EE. UU.): Educational Testing Service, 1992.

Meyer, A.; Rose, D. H. y Gordon, D. Universal design for learning: Theory and Practice. Wakefield, Massachusetts (EE. UU.): CAST Professional Publishing, 2014.

Meighan, R. "Restructuring Education so it works for kids & society". En Life Learning Magazine [en línea]. <a href="http://lifelearningmagazine.com/0412/restructuring-education.htm">http://lifelearningmagazine.com/0412/restructuring-education.htm</a>. Abril de 2017-01-15].

Mitra, S. Discurso en la Conferencia para directores de Colegios del Mundo del IB de 2011. Singapur. 13-16 de octubre de 2011.

Murphy, P. Learners, Learning & Assessment. Londres (Reino Unido): Paul Chapman Publishing en asociación con The Open University. 1999.

National Research Council. Assessing 21st Century Skills: Summary of a Workshop. Washington D. C. (EE. UU.): The National Academies Press, 2011.

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.

Newton, P. E. "We need to talk about Validity". Ponencia presentada en la reunión anual del National Council for Measurement in Education. Vancouver (Canadá), 2012.

Nuttall, D. L.; Backhouse, J. K. y Willmott, A. S. Comparability of standards between subjects. Londres (Reino Unido): Evans/Methuen Educational, 1974. Schools Council Examinations Bulletin, 29.

Ofqual. Inter-Subject Comparability: A Review of the Technical Literature: ISC Working Paper 2. Coventry (Reino Unido): Office of Qualifications and Examinations Regulation, 2015.

Peterson, A. D. C. New Techniques for the assessment of Pupils' Work. Estrasburgo (Francia): Consejo de Europa, 1971.

Peterson, A. D. C. Schools Across Frontiers: The Story of the International Baccalaureate and the United World Colleges. 2.ª ed. Chicago (EE. UU.): Open Court Publishing Company, 2003.

Popham, W. J. Criterion-referenced Measurement. Englewood Cliffs, Nueva Jersey (EE. UU.): Prentice Hall,

RAND Education. Teaching and Learning 21st Century Skills: Lessons from the Learning Sciences. Nueva York (EE. UU.): Asia Society, 2012. Disponible en línea en <a href="http://www.rand.org/pubs/external\_publications/">http://www.rand.org/pubs/external\_publications/</a> EP51105.html>. [Consulta: 2017-01-01].

Rao, K.; Currie-Rubin, R. y Logli, C. UDL and Inclusive Practices in IB Schools Worldwide. Wakefield, (EE. UU.): CAST Professional Learning. 2016.

Schleicher, A. "The case for 21st-century learning". En OCDE. OECD.org [en línea]. 2016. <a href="http://">http://</a> www.oecd.org/general/thecasefor21st-centurylearning.htm>. [Consulta: 2017-01-01].

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*. Boston (EE. UU.) y Dordrecht (Países Bajos): Kluwer, 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.

Smith, D. y Tomlinson, S. The School Effect. Londres (Reino Unido): Policy Studies Institute, 1989.

Snyder, B. R. The Hidden Curriculum. Cambridge, Massachusetts (EE. UU.): MIT Press, 1971.

Surgenor, P. Teaching Toolkit: Effect of Assessment on Learning. Dublín, (Irlanda): UCD Dublin, 2010.

Vygotsky, L. S. Thought and language. Nueva York (EE. UU.): Wiley, 1996.

Wiliam, D. "Validity, dependability and reliability in National Curriculum assessment". En *The Curriculum Journal*. 1993, vol. 4, n.º 3. Pp. 335-350.

En The Curriculum Journal. 1963, vol. 4, n.º 3. Pp. 335-350.

Winkley, J. y Cresswell, M. "Introduction to the concept of reliability". En Ofqual. *Ofqual's Reliability Compendium*. Capítulo 1. Coventry (Reino Unido): Office of Qualifications and Examinations Regulation, 2010.

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. Pp. 89-100.

Wood, R. Assessment and Testing: A survey of research. Cambridge (Reino Unido): Cambridge University Press, 1991.

Wood, D. How Children Think and Learn: The Social Contexts of Cognitive Development. 2.ª ed. Oxford (Reino Unido): Blackwell, 1998.

OCDE. "21st century skills and competences in OECD countries". En OECD.org [en línea]. Diciembre de 2009. <a href="https://www.oecd.org/edu/ceri/44254873.ppt">https://www.oecd.org/edu/ceri/44254873.ppt</a>>. [Consulta: 2017-01-31].



## Glosario

Término	Definición
Acceso inclusivo	Evaluación que tiene en cuenta las necesidades de todos los alumnos para que puedan 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 con necesidades específicas de acceso a la evaluación que le permitirán demostrar su nivel de conocimiento y capacidad de una forma más justa. No se han concebido para compensar una falta de aptitud.
Alumno	Persona que estudia un curso o un programa del IB. Los alumnos pasan a denominarse <i>alumnos matriculados</i> cuando forman parte del proceso de evaluación.
Alumno matriculado	Alumno inscrito para la evaluación.
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 se ofrecerá de manera general.
Autor del examen	Persona que crea las preguntas y el esquema de calificación asociado que se utilizará para las evaluaciones.
Banco de preguntas	Recopilación de preguntas proporcionadas junto con información acerca del tema 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 que demuestre ciertas cualidades.
Calificación	Descripción del logro de un alumno. Las calificaciones finales 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 y es coherente entre años y asignaturas.

Término	Definición
Carpetas electrónicas	Sistema o proceso mediante el cual los colegios cargan el material de examen o de trabajo de clase de los alumnos evaluado internamente 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	Se considerarán "circunstancias adversas" todas aquellas circunstancias ajenas a la voluntad del alumno matriculado que puedan perjudicar su desempeño durante la evaluación, incluidas 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 de los alumnos matriculados.  Las mismas circunstancias pueden afectar a un grupo de alumnos matriculados o a todos los alumnos matriculados de un colegio. Entre las circunstancias adversas no se incluyen las siguientes:  Deficiencias del colegio donde esté matriculado el alumno  El fracaso de los alumnos matriculados para mejorar su desempeño a pesar de habérsele autorizado adecuaciones inclusivas de acceso.
Circunstancias excepcionales	Circunstancias que comúnmente no se corresponden con las de otros alumnos 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 inclusivas de acceso.
Componente	Véase "Componente de evaluación".
Componente de evaluación	Un componente de evaluación consiste en una o varias tareas que, conjuntamente, forman parte de la evaluación general. Por ejemplo, una prueba de examen, una carpeta de trabajos, un proyecto o una tarea de investigación.
Concesión de calificaciones	En el proceso de concesión de calificaciones se decide cómo convertir las puntuaciones en calificaciones finales para garantizar que las calificaciones finales tengan el mismo significado, independientemente de la convocatoria en la que el alumno realice el examen.
Conducta improcedente	Toda acción (ya sea deliberada o involuntaria) de un alumno matriculado por la cual este u otro alumno matriculado sale o puede salir beneficiado injustamente en uno o varios componentes de la

Término	Definición
	evaluación. Las acciones que pudieran perjudicar a otros alumnos matriculados también se consideran conducta improcedente.
Conducta improcedente	Cualquier práctica contraria a los principios de probidad académica (por ejemplo, el plagio).
Consideraciones especiales	Cuando un alumno se ve afectado por circunstancias adversas, puede tener derecho a consideraciones especiales, siempre y cuando dichas consideraciones no lo ubiquen en una situación de ventaja respecto de otros alumnos matriculados. En tales casos, si el alumno se encuentra a uno o dos puntos de la banda de la calificación superior para la disciplina o las disciplinas en cuestión, la calificación se aumentará. Esta es la única modificación que podrá hacerse en el caso de alumnos afectados por circunstancias adversas. Si la puntuación del alumno no se encuentra dentro del rango requerido, no se realizará ningún ajuste.
Consulta sobre los resultados	Revisión de los niveles (puntuación) llevada a cabo a petición del colegio.
Convocatoria	Véase "Convocatoria de exámenes".
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 electrónica	Proceso mediante el cual los examinadores corrigen el material de examen directamente en la pantalla de la computadora.
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 de los docentes (el currículo enseñado) y a lo que los alumnos aprenden (el currículo evaluado).
Criterios de evaluación	Criterios en relación con los cuales se mide el desempeño de un alumno.
Criterios holísticos	Enfoque de evaluación que considera el trabajo del alumno como un resultado único, en lugar de analizar distintos elementos individuales de dicho trabajo (como la comunicación, el conocimiento de la materia, la calidad del argumento, etc.).
Curso	Número prescrito de clases, lecciones u horas lectivas en un período de estudio definido. Los colegios organizan la enseñanza y el aprendizaje de las asignaturas por medio de cursos disciplinarios e interdisciplinarios.
Descriptores de calificaciones finales	Explicación de las cualidades que se esperan en los alumnos para alcanzar cada calificación. Un

Término	Definición
	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.
Descriptores de nivel de logro	Describen las características del trabajo del alumno que cabría esperar en cada nivel de logro.
Diferenciación (en la enseñanza y el aprendizaje)	Modificación de las estrategias de enseñanza para atender las diversas necesidades de los alumnos por medio de distintos contenidos, procesos y productos.
Diferenciación (en la evaluación)	Distinción entre alumnos 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 "asignatura" en el PAI y el PD para fines prácticos de evaluación). Algunos grupos de asignaturas y asignaturas del PAI pueden comprender varias disciplinas. Por ejemplo, el grupo de asignaturas llamado Artes comprende disciplinas tales como Arte Visual, Teatro, Música, Medios de Comunicación y Danza. El curso integrado de Ciencias comprende tres disciplinas: Biología, Física y Química.
Diseño universal de evaluación	El concepto de diseño universal de evaluación es que todas las evaluaciones se deben desarrollar teniendo en cuenta la variedad de necesidades que pueden tener los alumnos, en vez de tratar a algunos alumnos de manera distinta. Esto forma parte del compromiso del IB con el diseño universal para el aprendizaje (DUA).
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 enseñanza y el aprendizaje cambian según la forma en que se evalúe al alumno.
Elemento de la pregunta o subpregunta	La unidad más pequeña de una pregunta o tarea de evaluación. En la corrección, cada elemento recibe una cantidad de puntos. Un elemento puede ser una pregunta completa o partes de una pregunta.
Enfoque basado en criterios	Comparación del desempeño del alumno con descripciones predefinidas de logro (criterios) para realizar la corrección.



Término	Definición
Enfoque flexible basado en criterios	Si el logro del alumno se compara con descripciones predefinidas de logro (los criterios) y con el rendimiento de anteriores grupos de alumnos, esto se denomina enfoque flexible basado en criterios. Este es el enfoque que el IB adopta para mantener los estándares.
Enfoque normativo	Cuando el logro se determina al comparar el desempeño del alumno con el del conjunto de alumnos para quienes se diseñó la evaluación.
Envío	Trabajo final que el alumno (o el colegio en representación del alumno) envía al IB.
Equiparabilidad	El grado en que un resultado concreto puede considerarse idéntico a otro. Se aplica generalmente entre años (por ejemplo, si una calificación de 8 este año es idéntica a una calificación de 8 el año pasado) o entre asignaturas (por ejemplo, si una calificación de 5 en Matemáticas es idéntica a una calificación de 5 en Artes).
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.
Estándar	Nivel de desempeño que se espera para lograr una puntuación, una calificación o un resultado concreto.
Estandarización	Proceso colaborativo en el que los moderadores o examinadores acuerdan un estándar de evaluación común.
Estandarización interna	Proceso mediante el cual todos los profesores de un colegio que imparten una asignatura concreta se aseguran de que están aplicando el mismo estándar en la evaluación.
Estrategia de evaluación	Método o enfoque empleado para obtener información sobre el aprendizaje de los alumnos (por ejemplo, observación, tareas abiertas o respuestas seleccionadas).
Evaluación	Recopilación de indicios que permiten valorar la enseñanza y el aprendizaje.
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 depende del nivel de logro alcanzado por todo el grupo de alumnos, sino que es fijo.
Evaluación electrónica	Evaluación realizada a través de una computadora o un dispositivo similar.

Término	Definición
Evaluación externa	Evaluación establecida y corregida por el IB, no por el profesor del alumno.
Evaluación final	Evaluación sumativa del trabajo del alumno al término del período de estudio.
Evaluación formativa	Evaluación continua del trabajo del alumno 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 los profesores en el colegio.
Evaluación sumativa	Evaluación destinada a determinar la competencia o el nivel de logro de un alumno, y que normalmente tiene lugar al completar un programa de estudios o una unidad de trabajo.
Examen	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 los alumnos deben realizar en condiciones controladas y con un límite de tiempo.
Examen con las respuestas del alumno o tarea de evaluación enviada	Respuestas del alumno a la prueba de examen. También puede tratarse de un trabajo que el alumno haya entregado para la evaluación.
Examen de aptitud	Un ejemplo de trabajo de un alumno que selecciona el examinador principal para realizar una comprobación formal de que los examinadores han comprendido el estándar requerido de corrección antes de que se les permita corregir en condiciones reales.
Examen de control	Examen corregido por el examinador principal y que se incluye aleatoriamente en un lote de exámenes asignado a un examinador para su corrección.  Aparentemente es igual que cualquier otro examen, por lo que el examinador no sabrá que se trata de un examen de control. Los puntos que el examinador concede al examen de control se comparan con los concedidos por el examinador principal, con un margen de tolerancia, para comprobar si el examinador está aplicando el estándar de corrección prefijado. Los exámenes de control de moderación



Término	Definición
	de la toma dinámica de muestras se utilizan del mismo modo como parte del proceso de moderación.
Examen de práctica	Ejemplos de trabajos de alumnos que se identifican y corrigen durante la estandarización y que, a continuación, se entregan a los examinadores para que conozcan el estándar que deben aplicar en su corrección.
Examen electrónico	Respuestas del alumno a una tarea de evaluación electrónica.
Examen en pantalla	Examen multimedia formal, con límite de tiempo y elaborado externamente que incluye una serie de tareas relacionadas con la asignatura y diseñadas para responderse en un entorno de examen seguro.
Examinador	Persona que concede puntuaciones a las tareas de evaluación externa de un alumno.
Examinador jefe	El examinador 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 (por ejemplo, Ciencias).
Examinador principal	El examinador principal es el responsable de dirigir la evaluación de un componente. Esta persona establece el estándar de corrección y normalmente es también uno de los autores de la evaluación.  En el PAI, la función del examinador principal difiere ligeramente con respecto a otros sistemas de evaluación. El examinador principal es el 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 de los equipos de examinadores.
Examinador supervisor	Examinador con experiencia que ayuda al examinador principal.
Factor de moderación	Ajuste aritmético aplicado a la puntuación total concedida por el evaluador para adecuarla al estándar de evaluación común.
Fiabilidad	La medida en que el alumno logrará el mismo resultado en cada evaluación de su trabajo. Puede hacer referencia a la fiabilidad entre examinadores (es decir, si todos conceden el mismo resultado al alumno) o a la fiabilidad de un único examinador (es decir, si concede el mismo resultado en cada corrección del trabajo del alumno).
Grupo de preguntas (QIG, por sus siglas en inglés)	Se considera un grupo a una o varias preguntas de un cuestionario de examen que estén relacionadas entre sí. Los examinadores deben corregir los QIG individuales, en lugar de exámenes completos. Este

Término	Definición
	enfoque ofrece unos resultados más fiables que la corrección de exámenes completos.
Herramienta de evaluación	Método para obtener información sobre el desempeño y la comprensión de un alumno.
Herramienta de familiarización	Simulación genérica de un examen que pueden realizar los alumnos para aprender a utilizar las herramientas y el entorno del examen en pantalla.
IBIS	Sistema que permite a los coordinadores de los colegios realizar procedimientos administrativos y obtener noticias e información del IB por medio de un servidor web al que se accede con una contraseña.
Jefe de equipo	Examinador que dirige un equipo de examinadores.
Lengua de respuesta	Lengua que el alumno 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, estas lenguas son el español, el francés y el inglés.
Límite de calificación	Punto en el que el nivel de logro de un alumno 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	El grado en el que la evaluación y cada una de las tareas suponen una carga para el alumno 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 nota definitiva concedida por el examinador principal que el IB considera lo suficientemente próxima como para demostrar que el examinador está aplicando el estándar de corrección correcto. Los márgenes de tolerancia son necesarios, ya que la corrección no deja de ser una cuestión de opinión e incluso los examinadores 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.

Término	Definición
Material de ayuda al profesor	Información adicional para ayudar a los profesores a entender los requisitos del curso del IB. Su finalidad es proporcionar ayuda práctica para la comprensión e implementación de la teoría que se encuentra en las guías de las asignaturas.
Matriculación de alumnos	Proceso llevado a cabo por el coordinador del colegio para inscribir a los alumnos para la evaluación del IB.
Modelo de calidad	Enfoque adoptado por el IB para garantizar que los alumnos reciben el resultado correcto de la evaluación. El examinador principal define el estándar de respuesta correcto para cada pregunta y todos los examinadores deben reproducirlo. En el caso de los trabajos evaluados externamente, el proceso implica proporcionar orientación a los examinadores por medio de la estandarización, comprobar su comprensión del estándar mediante exámenes de aptitud y supervisar regularmente su corrección por medio de exámenes de control.
Moderación	Proceso para garantizar la aplicación de un estándar de evaluación común por medio de la revisión de muestras de trabajos evaluados y el ajuste de las calificaciones de los evaluadores, según sea necesario.
Moderación externa	Véase "Moderación".
Muestra para moderación	Muestra de trabajos de alumnos que se envía al IB para asegurarse de que está corregida de acuerdo con el estándar requerido.
Necesidades específicas de acceso a la evaluación	Un alumno con necesidades específicas de acceso a la evaluación es aquel que necesita cambios en las condiciones de evaluación para poder demostrar su nivel de conocimiento y capacidad.
Nivel de logro	Valor numérico que se otorga al trabajo del alumno 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	El grado en el que la evaluación permite comprobar realmente los conocimientos y habilidades en cuestión. Comprobar las habilidades prácticas de un alumno por medio de un examen escrito sería un ejemplo de poca pertinencia del constructo.

Término	Definición
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.
Predictibilidad	El término predictibilidad designa la capacidad de estimar qué sucederá o cuándo sucederá algo. En el ámbito de la evaluación, esto se refiere a la capacidad de los colegios para prever qué preguntas aparecerán en los exámenes y cuándo. Es fundamental que las prácticas del IB en materia de evaluación tengan una buena predictibilidad, ya que, al cumplir este principio, el IB muestra que respeta los requisitos de sus constructos tal como se publican para los profesores, lo cual da pie a una oportunidad de evaluación "justa" en lo referido al cumplimiento curricular (por ejemplo, lo que el IB haya dicho que se va a evaluar, se evaluará).
Pregunta	Tarea o actividad utilizada para que un alumno demuestre su competencia en una asignatura.
Pregunta de opción múltiple	Pregunta en la que el alumno debe seleccionar la respuesta correcta de una lista de posibles respuestas que se proporcionan.
Probidad académica	Conjunto de valores y habilidades que promueven la integridad personal y las buenas prácticas en la enseñanza, el aprendizaje y la evaluación.
Procedimiento de ausencia de calificación	Mecanismo para otorgar una calificación a los alumnos en los casos en los que el IB no puede acceder a una calificación justa o precisa basada en el trabajo que ha completado el alumno. Está pensado para aquellas circunstancias en las que la ausencia de pruebas se debe a alguna acción del IB o de terceros (a excepción del colegio), a consecuencia de la cual no sería razonable exigir al alumno que completara la evaluación en otra ocasión.
Procedimiento de ausencia de puntuación	Mecanismo para otorgar una puntuación a los alumnos en los casos en los que el IB no puede acceder a una puntuación justa o precisa basada en el trabajo que ha completado el alumno. Está pensado para aquellas circunstancias en las que la ausencia de pruebas se debe a alguna acción del IB o de terceros (a excepción del colegio), a consecuencia de la cual no sería razonable exigir al alumno que completara la evaluación en otra ocasión.
Prueba	Véase "Prueba o cuestionario de examen".
Prueba modificada	Cambios efectuados en una evaluación para permitir que un alumno con necesidades específicas pueda realizarla en las mismas condiciones que los



Término	Definición
	alumnos que no presentan 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.
Prueba o cuestionario de examen	Conjunto de tareas y preguntas que se pide a un alumno que complete. En determinadas circunstancias, este término puede referirse a un examen que se responde en pantalla o a un examen que se responde en papel.
Publicación de resultados	Proceso mediante el cual los alumnos reciben las calificaciones del IB correspondientes a sus evaluaciones.
Puntuación	Valor que refleja la calidad de la respuesta del alumno a una pregunta específica planteada.
Puntuación definitiva	Puntuación que el examinador principal concede a un trabajo concreto del alumno. Representa la corrección que todos los demás examinadores deberían intentar replicar. (Véase también "Modelo de calidad").
Repetidor	Categoría de matriculación para alumnos que se presentan nuevamente a uno o varios exámenes con la esperanza 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.
Respuesta atípica	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 conducta improcedente.
Respuesta de evaluación	Término utilizado para describir todo el material elaborado por un alumno en respuesta al material de evaluación.
Resultados de los cursos	Documento principal en el que se informa de los resultados que ha obtenido el alumno. Muestra todas las disciplinas estudiadas por el alumno y la calificación final que ha logrado (1-7). También muestra la calificación recibida en los componentes troncales, la evaluación interdisciplinaria y si se ha completado el requisito de servicio comunitario del colegio.  Por último, el documento incluye el nombre del alumno, 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

Término	Definición
	(además de un duplicado, si procede). El documento de resultados solo hace constar los logros positivos.
Reunión de estandarización	Reunión de los examinadores principales para describir el estándar de corrección que debe aplicarse y definir los exámenes de control.
Revisión de la corrección del examinador	Proceso de revisión de la corrección de las respuestas electrónicas asignadas a un examinador si se detecta que la corrección no se ajusta al estándar esperado o difiere considerablemente de este. Con frecuencia, esto se debe a un error en la moderación.
Sesgo	Se considera que existe sesgo cuando un grupo definido (por ejemplo, racial, étnico, o de un género) tiene, en una pregunta o tarea específica, un desempeño distinto al promedio por algún motivo que no sea su capacidad en el aspecto que se evalúa.
Supervisor de exámenes	Persona que supervisa y controla el entorno de examen.
Tarea de evaluación	Actividad o serie de actividades que realizan los alumnos 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. Aplica el modelo de calidad de "tolerancia" a las calificaciones de los profesores y los examinadores. En el caso de los profesores, si la muestra inicial se encuentra dentro del margen de tolerancia, no se aplicará ningún factor de moderación. Además, los moderadores (examinadores) reciben los trabajos de los alumnos de manera individual, lo que permite incluir "exámenes de control" para mantener un estándar coherente. El proceso también permite asignar a los examinadores los exámenes adicionales que sean necesarios si hay indicios de que la corrección del profesor no se ajusta al estándar general.
Totalización	Proceso de combinar las puntuaciones obtenidas para obtener un resultado final.
Trabajo evaluado externamente	Trabajo evaluado o corregido íntegramente por el IB.
Trabajo evaluado internamente	Trabajo evaluado (corregido) por los profesores de los alumnos. Con el fin de realizar la moderación, el IB pide muestras del trabajo evaluado internamente.



Término	Definición
Validez	Término general 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 en relación con un criterio de evaluación determinado.
Verificación de la autoría original	Proceso de demostración y prueba de que el trabajo ha sido llevado a cabo por el propio alumno. Ejemplos pueden ser las firmas del profesor y el alumno que indican el origen de la respuesta del alumno.