

Principios y prácticas de evaluación del IB: guía sobre la evaluación para examinadores jefe y examinadores principales

Introducción

Esta guía está pensada para proporcionar información general sobre los procesos de evaluación pertinentes para examinadores jefe y examinadores principales. Estos dos tipos de examinadores normalmente forman parte de todas las fases del proceso de evaluación. Si bien es posible que la mayoría de la información que se presenta aquí le resulte familiar, esperamos que al presentar las responsabilidades de cada tipo de examinador pueda entender la importancia de cada una de las fases de la evaluación.

¿Aúna las distintas responsabilidades que tiene?

Las evaluaciones sumativas deben proporcionar a los alumnos resultados *significativos*, *justos* y *fiabes*:

- **Significativos** porque las evaluaciones deben medir aquello que se pretendía ponderar, tal como se indicara en la guía de la asignatura.
- **Justos** porque las evaluaciones no deben favorecer a ningún grupo concreto de alumnos.
- **Fiabes** porque los alumnos deben recibir los mismos resultados independientemente de quién corrija sus evaluaciones o en qué día lo hagan.

Una evaluación que no sea significativa, que no resulte justa o que no dé resultados fiabes no será adecuada para utilizarla como una evaluación sumativa de la calidad que espera el IB.

La función del examinador jefe es supervisar y responsabilizarse de la evaluación de una asignatura, siempre teniendo en mente el propósito de la evaluación. Esta función también implica asegurarse de que se mantenga el mismo estándar de un año al otro.

El examinador principal se encarga del estándar de corrección de un componente de una asignatura. Se espera que todos los examinadores que corrijan un componente lo hagan en consonancia con el estándar del examinador principal, ya que así se garantiza que todos los alumnos reciban los resultados adecuados; es decir, que el resultado de un alumno no dependa del examinador que haya corregido su trabajo.

El siguiente diagrama muestra las fases del ciclo de evaluación, un flujo circular pensado para mejorar continuamente. Esto implica un esfuerzo constante para lograr que las evaluaciones sean más significativas, más fiabes y más justas.

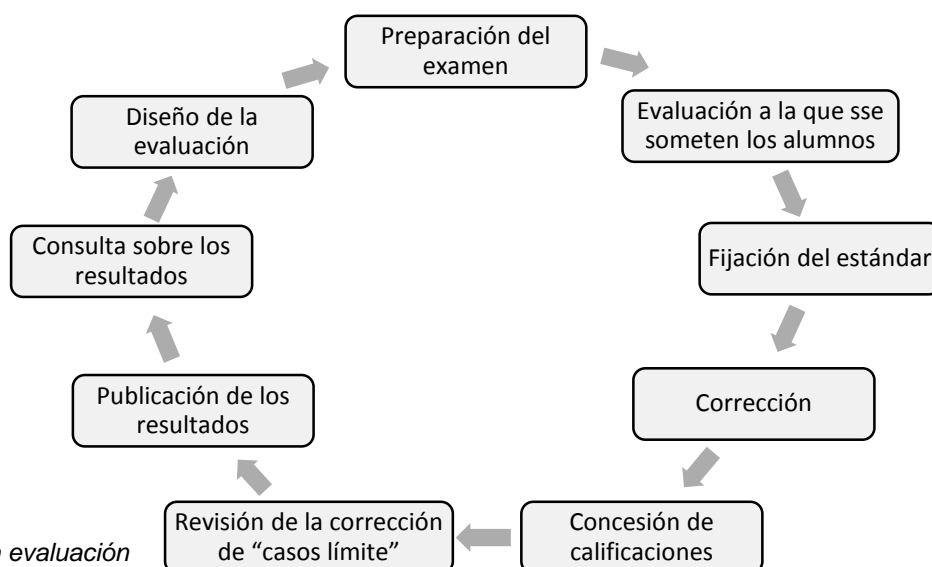


Figura 1. Ciclo de la evaluación

Qué es importante al preparar las evaluaciones y los esquemas de calificación

El examinador jefe y los examinadores principales, con el apoyo de otros examinadores experimentados, son los responsables de preparar los exámenes de cada convocatoria, un trabajo que puede empezar entre 18 meses y 2 años antes de que los alumnos se sometan a los exámenes. Se debe prestar atención para que todos los exámenes y evaluaciones que conforman el conjunto exigido para la asignatura sean pruebas equilibradas y que en conjunto cubran todos los objetivos. El examinador supervisor puede redactar todo el examen o bien componerlo a partir de preguntas que aporten otros examinadores.

La predictibilidad es una cuestión difícil a la hora de crear evaluaciones. En todos los exámenes, los alumnos deben afrontar preguntas o tareas que sean de cierta manera distintas de lo que hayan hecho anteriormente (para evitar los efectos negativos de una enseñanza centrada en las pruebas de evaluación), si bien también deben ser equiparables a ejemplos anteriores (para que los alumnos se puedan preparar con eficacia). El IB procura evaluar habilidades cognitivas de orden superior; por tanto, debe haber preguntas que pidan a los alumnos solucionar un problema o pensar de manera creativa para aplicar sus conocimientos en un contexto nuevo, en vez de limitarse a mostrar habilidades que tengan bien practicadas o a repetir sus conocimientos.

Las preguntas y los esquemas de calificación se deben preparar conjuntamente para garantizar que van en consonancia con lo que pretenden evaluar. Se debe dedicar el mismo esfuerzo a la preparación de un esquema de calificación que a la de las preguntas durante el proceso de creación de exámenes. También se deben preparar notas explicativas para la corrección incluso para los exámenes con preguntas o tareas abiertas que se corrijan según los mismos criterios de evaluación en cada convocatoria; estas notas ofrecen orientación sobre cómo aplicar dichos criterios en el contexto de cada pregunta. Los esquemas de calificación resultan tan importantes para la integridad del proceso de evaluación como los propios exámenes. El esquema de calificación es mucho más que un simple conjunto de ejemplos de respuestas: ofrece orientación sobre cómo puntuar enfoques comunes y alternativos que los alumnos pueden adoptar cuando responden a una pregunta, y también da información sobre cómo abordar equivocaciones o errores habituales de los alumnos.

Los esquemas de calificación y las notas para la corrección deben ser claros y reducir al mínimo la posibilidad de que los examinadores aborden la corrección de manera distinta los unos de los otros, ya que representan la herramienta principal para comunicar a los examinadores lo que se espera de ellos.

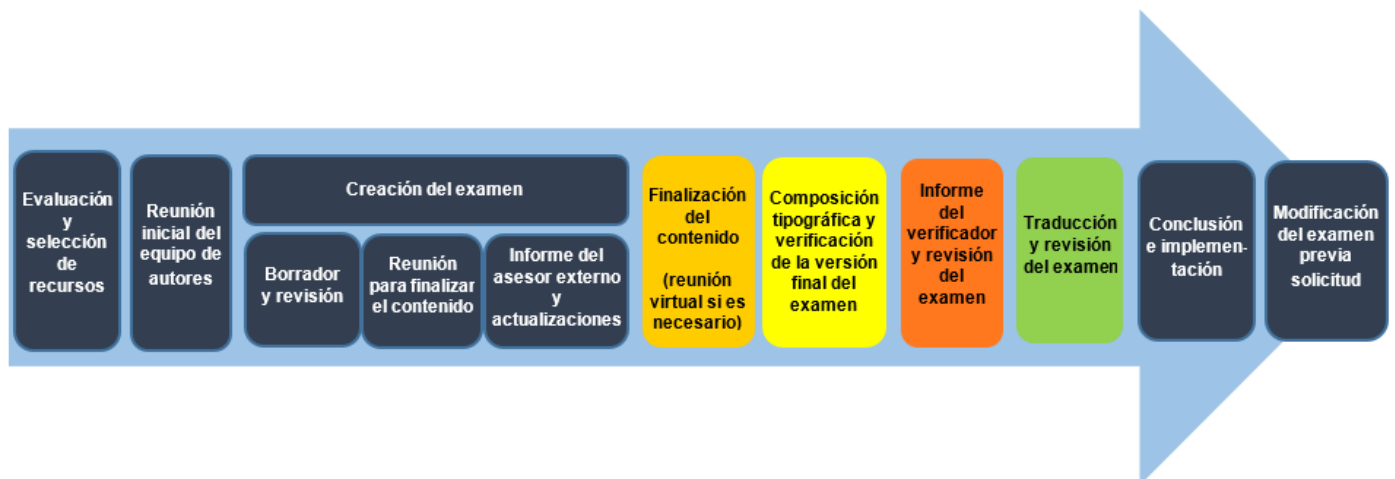


Figura 2. Preparación de los exámenes.

Cómo puede comunicar su estándar a los examinadores

La estandarización no consiste en una única reunión, sino que es un período de tiempo con varios objetivos:

- Que el examinador principal determine el estándar para la evaluación
- Poner a prueba y perfeccionar el esquema de calificación
- Elaborar exámenes corregidos definitivos (de práctica, de aptitud y de control)

- Transmitir lo aprendido a todos los examinadores
- Confirmar que los examinadores entienden el estándar

Resulta importante comprender los propósitos de los tres tipos de exámenes corregidos definitivos. El propósito de los exámenes de práctica es ayudar a los examinadores a aprender el estándar de corrección, mientras que los exámenes de aptitud sirven para demostrar o probar que los examinadores son capaces de corregir siguiendo el estándar adecuado. Por su parte, el propósito de los exámenes de control es demostrar o probar que los examinadores siguen corrigiendo de acuerdo con el estándar pertinente.

Los exámenes seleccionados deben mostrar una calidad clara, ya que su finalidad es comprobar si los examinadores aplican los esquemas de calificación o los criterios de evaluación adecuadamente, y si corrigen de acuerdo con el estándar de corrección del examinador principal; el objetivo no es “sorprender al examinador con la guardia baja”. Es importante que los exámenes seleccionados contengan respuestas y puntuaciones variadas para comprobar que los examinadores comprenden exhaustivamente los esquemas de calificación y los criterios de evaluación, algo que resulta esencial para poder poner en práctica el estándar de corrección del examinador principal.

Para saber si se ha desarrollado con éxito la fase de fijar el estándar de corrección, hay que ver si los examinadores comprenden bien dicho estándar y si el IB tiene un conjunto de exámenes corregidos definitivos (exámenes de práctica, de aptitud y de control) para respaldar a los examinadores en su aprendizaje y para que demuestren que dominan el estándar de corrección.

Cómo puede saber si los examinadores comprenden su estándar de corrección

A fin de garantizar unos resultados justos para todos los alumnos, resulta de vital importancia la fiabilidad en la corrección, es decir, la capacidad de que el proceso de evaluación otorgue casi la misma puntuación a un trabajo sin importar quién sea el examinador que lo corrija y en qué momento lo haga. En cada componente, el examinador principal fija el estándar de corrección y debe comunicarlo con claridad para que todos los examinadores puedan corregir en consonancia a dicho estándar. Si no se sigue este estándar, es posible que los resultados que obtengan los alumnos no sean fiables, y eso significa que la evaluación no cumplirá su propósito.

En el caso de los componentes que se corrigen electrónicamente, para garantizar la calidad de la corrección se utiliza un *sistema de control*. Durante el proceso de control, los examinadores corrigen exámenes que ya han corregido previamente el examinador principal y el equipo de examinadores supervisores, a fin de poder comparar los dos grupos de puntuaciones. Si las diferencias entre los dos grupos de puntuaciones se encuentran dentro de un margen de tolerancia predeterminado, se considera que la corrección es aceptable. Se aplica un margen de tolerancia porque se acepta que puede haber diferencias legítimas en las puntuaciones que distintos examinadores otorguen a un mismo trabajo; sin embargo, es importante reducir al mínimo estas diferencias para que afecten de la menor manera posible a los resultados de los alumnos.

Estos márgenes de tolerancia variarán para las distintas tareas, ya que se tiene en cuenta que es más fácil coincidir en la corrección de aspectos objetivos que en la de aspectos subjetivos. En muchos casos, resulta razonable esperar que todos los examinadores otorguen la misma puntuación; un ejemplo sería aquellos casos en los que está claro si la respuesta es correcta o incorrecta. En preguntas que requieren una respuesta tipo ensayo, los examinadores deben coincidir en la calidad general de la respuesta, pero al corregirla que haya uno o dos puntos de diferencia.

Para garantizar que el trabajo de los alumnos se corrija de manera precisa y reducir al mínimo la necesidad de revisar la corrección al final del período de corrección, los examinadores deben corregir una cantidad determinada de exámenes de aptitud para demostrar que pueden corregir en consonancia con el estándar establecido antes de pasar a corregir exámenes en condiciones reales. Incluso cuando ya están corrigiendo en condiciones reales, a los examinadores se les controla para comprobar si corrigen según el estándar establecido. Para ello, se introducen exámenes de control entre las pruebas que deben corregir los examinadores. Los examinadores no deben saber qué exámenes son de control, con el fin de que los corrijan como harían con los demás.

Las puntuaciones que otorga el examinador se comparan siempre con las del examinador principal para comprobar si los alumnos reciben resultados fiables. El IB no adopta ninguna posición sobre cuál debe ser la puntuación correcta.

Cómo proporcionar comentarios a los examinadores

Si la puntuación de un examinador es muy distinta a la puntuación definitiva o acordada durante la corrección en condiciones reales, es importante que el examinador reciba inmediatamente comentarios sobre cómo se han otorgado las puntuaciones y por qué. Al examinador se le entrega una copia del examen de control en el que puede ver las puntuaciones definitivas y las anotaciones detalladas del equipo supervisor. Si, a pesar de esta información, el examinador continúa corrigiendo fuera del margen de tolerancia, no se le permitirá continuar con su labor hasta que un examinador supervisor o un jefe de equipo se ponga en contacto con él para conversar sobre su corrección.

El objetivo de esta conversación es que los jefes de equipo ofrezcan comentarios de calidad que permitan a los examinadores entender cómo deben realizar sus correcciones. Los examinadores deben 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 utilizarse para garantizar que los examinadores cumplan con el estándar de corrección que fijara el examinador principal.

En el momento en que el jefe de equipo esté de acuerdo con que el examinador puede volver a corregir, le permitirá el acceso a más exámenes. Sin embargo, si el examinador continúa corrigiendo exámenes de control de manera inadecuada, se le volverá a retirar de la corrección. En esta ocasión, la información se comunicará al personal del IB, que deberá decidir si retirar de la corrección al examinador de manera permanente para evitar que los alumnos reciban puntuaciones incorrectas.

La proporción de examinadores que corrigen los exámenes de control dentro del margen de tolerancia indica el nivel de fiabilidad de la corrección de un componente. Si solo una proporción relativamente pequeña de examinadores corrige los exámenes de control dentro del margen de tolerancia, el examinador jefe y el examinador principal deben determinar la causa de este bajo nivel de fiabilidad de la corrección y abordar esta causa para futuras convocatorias. Por ejemplo, si el esquema de calificación no resultó lo suficientemente claro para explicar cómo se deben otorgar las puntuaciones, se deben revisar los esquemas de calificación futuros para ofrecer una orientación más clara. Si la tarea de evaluación no permite una corrección fiable, se debe revisar el currículo para permitir cambiar la tarea o sustituirla por otra.

A la hora de fijar la puntuación definitiva en los exámenes de control se debe ir con mucho cuidado, no solo por cómo se usan en el proceso de garantía de la calidad, sino también porque se espera que los examinadores utilicen las puntuaciones y las anotaciones para perfeccionar su corrección. Algunos posibles efectos de un error en un examen corregido definitivamente son: que a un examinador se le impida injustamente seguir corrigiendo, que no se impida a los examinadores seguir corrigiendo (lo que provoca que los alumnos reciban resultados injustos), que se ofrezca orientación incorrecta a los examinadores, y que se dé una puntuación incorrecta a los alumnos cuyos trabajos se hayan corregido definitivamente.

Grupos de preguntas

Algunos exámenes se corregirán por grupos de preguntas (*question item group*, QIG), que pueden ser parte de la respuesta de un alumno o bien toda la respuesta. Si el examen se divide en distintas partes, habrá varios examinadores que corrijan las distintas secciones. Al dividir un examen en partes, por grupos de preguntas separados, los examinadores pueden elegir qué grupo de preguntas quieren corregir según sus conocimientos y preferencias. Para supervisar cómo los examinadores corrigen los grupos de preguntas, se utiliza el sistema de exámenes de control.

Moderación de la evaluación interna

La moderación es la herramienta principal para garantizar la fiabilidad de la corrección que los profesores realizan de los trabajos de evaluación interna. A diferencia de la corrección, la moderación está pensada para comprobar que los profesores apliquen el esquema de calificación correctamente, pero da por hecho que los profesores aplican su criterio de manera coherente.

Qué sucede si hay pruebas contradictorias al conceder calificaciones

Desde el punto de vista del alumno, la reunión de evaluación representa la culminación del proceso de evaluación. La concesión de calificaciones aporta sentido a las puntuaciones que han recibido los alumnos. Cabe destacar que las puntuaciones pueden variar de una convocatoria a otra debido a que hay preguntas distintas, pero se debe mantener el estándar de calificación utilizado en convocatorias anteriores. Por ejemplo, una calificación de 4 en la convocatoria vigente representa el mismo estándar que una calificación de 4 en la convocatoria anterior, y en la anterior a esa. A fin de mantener

la coherencia, el equipo de concesión de calificaciones revisa los exámenes de convocatorias anteriores que se encuentran entre una banda de puntuación y otra.

La primera tarea que se hace en la reunión de evaluación es reflexionar sobre cómo ha funcionado cada componente. Todos los participantes se habrán implicado activamente en la corrección de un componente como mínimo, y la mayoría habrán contribuido a la redacción de los cuestionarios de examen. Además de su propia experiencia, los examinadores supervisores deben revisar los comentarios que los profesores hayan enviado formalmente sobre los cuestionarios de examen, los informes de los examinadores sobre sus percepciones acerca de los exámenes, y los datos estadísticos obtenidos de la corrección de los exámenes. Esta información proporciona un contexto importante para ayudar a acordar unos límites de calificación adecuados.

Posteriormente, el equipo debe plantearse cada componente para el cual hay que fijar nuevos límites en cada convocatoria. Normalmente, el cambio en las puntuaciones límite es leve, ya que se debe procurar que cada nueva versión de un examen esté al mismo nivel de dificultad general que la de la convocatoria anterior. El examinador jefe tiene la responsabilidad de tener en cuenta todas las pruebas disponibles (los comentarios sobre el funcionamiento del examen, la información estadística y las valoraciones del equipo supervisor) para decidir qué puntuación límite conserva mejor el estándar.

Los límites para los componentes de evaluación interna y para los componentes de corrección externa que no sean un examen no varían necesariamente en cada convocatoria, porque los requisitos de la tarea son los mismos a no ser que haya cambiado el curso. En esos casos, los límites se suelen fijar una sola vez al inicio del curso, aunque se pueden revisar en cada convocatoria.

Cuando se considera que los resultados finales son justos y correctos, el equipo de examinadores supervisores y otros examinadores con experiencia pueden dar respuesta a las cuestiones que queden pendientes en relación con la fiabilidad de la corrección. La revisión de la corrección se centrará principalmente en los alumnos que se considere que estén en "casos límite". Por lo general, estos son alumnos cuyo trabajo corrigieron examinadores cuyas correcciones se identificaron como potencialmente problemáticas, o bien alumnos cuya calificación final se encuentra dos calificaciones o más por debajo de lo previsto o que están a dos puntos porcentuales de alcanzar una calificación mejor. Como en la corrección hay un error de medición, en esos casos que se encuentran al límite se debe confirmar la precisión de la corrección. Por ese motivo, los correctores más fiables deben ser los que revisen el trabajo de los alumnos que estén en "casos límite".

Todas las discusiones y las decisiones de la reunión deben quedar registradas en un informe sobre la concesión de calificaciones, y se deben usar para ayudar a fijar las evaluaciones para la convocatoria siguiente.

Envío de comentarios a los colegios

El objetivo del proceso de evaluación no es dar a los colegios asesoramiento sobre la enseñanza que imparten, sino conceder calificaciones válidas a los alumnos. Sin embargo, el IB está comprometido a ofrecer apoyo a los colegios para mejorar los estándares. No resulta adecuado destinar recursos para proporcionar comentarios individuales; los informes generales de las asignaturas deben ofrecer información completa sobre el desempeño general de la convocatoria. Con esto se consigue un doble beneficio: fortalecer el apoyo que la evaluación sumativa aporta a la enseñanza en clase, y aclarar el funcionamiento del sistema de evaluación para los colegios, los profesores y los alumnos que lo utilizan.

Cuando se acaba una convocatoria de exámenes, los colegios pueden adquirir los exámenes y sus respectivos esquemas de calificación. Como examinador principal o examinador jefe, usted tiene la responsabilidad de redactar el informe general de la asignatura utilizando la información que le proporcione el equipo de examinadores. El informe general de la asignatura cubre todos los aspectos generales del desempeño de los alumnos en cada componente, destaca en qué partes los alumnos tuvieron un buen desempeño y en cuáles mostraron menos capacidad, y ofrece recomendaciones para mejorar la preparación de los alumnos.

El Comité de la evaluación final

El Comité de la evaluación final se reúne justo después celebrarse todas las reuniones de concesión de calificaciones y justo antes de la publicación de los resultados. Este comité concede formalmente los diplomas y los certificados a los alumnos que cumplen los requisitos. La reunión la dirige el presidente de la Junta de examinadores y participan en ella una cantidad reducida de examinadores jefe y personal sénior del IB. También se invita a un observador de un colegio.

El Comité de la evaluación final cuenta con un subcomité de probidad académica que establece qué acciones se deben tomar en caso de una presunta conducta impropia. El subcomité incluye a examinadores supervisores, profesores sénior de Colegios del Mundo del IB, y personal del IB. Un equipo formado al menos por un representante de cada uno de estos grupos, con un mínimo de tres personas, es el encargado de tomar todas las decisiones.

El comité también tiene en cuenta las recomendaciones para nuevas políticas que le llegan por parte de personal del IB sobre las adecuaciones de la evaluación que hay que hacer para los alumnos con necesidades educativas especiales. Además, el comité estudia los casos en que los colegios hagan una mala administración y no respeten los plazos o los procedimientos. Si hay casos graves que impliquen una amenaza importante a la seguridad y la integridad de los exámenes, o si hay una mala administración reiterada, se puede retirar la autorización al colegio.

¿Se es más generoso en las consultas sobre los resultados?

Los objetivos del servicio de consulta sobre los resultados son permitir que los colegios y los alumnos indiquen sus preocupaciones sobre la corrección de las evaluaciones, y ofrecer transparencia en lo referido a la concesión de puntuaciones. Este servicio no está pensado para que los alumnos obtengan un mejor resultado; por eso, los examinadores jefe y los examinadores principales deben liderar a sus equipos de examinadores de tal modo que se garantice que en la consulta sobre los resultados se sigan exactamente los mismos estándares que en el período de corrección inicial. Se debe tratar de manera justa a todos los alumnos, tanto si han solicitado una consulta sobre los resultados como si no lo han hecho. Una buena pregunta que se puede plantear es si el trabajo cuya corrección está revisando tiene la calidad suficiente para merecer una calificación más alta.

Si durante el proceso de consulta sobre los resultados se identifican problemas de corrección (es decir, si se cambian las puntuaciones), como examinador jefe o examinador principal debe plantearse en qué manera se pueden minimizar estos desacuerdos en el futuro y cómo se puede aumentar la fiabilidad de la corrección para garantizar que todos los alumnos obtengan resultados correctos.

Cómo mejorar las evaluaciones para la próxima vez

Es importante que aprendamos de nuestras experiencias durante el ciclo de evaluación para que podamos mejorar tanto los exámenes como los modelos que utilizamos para evaluar a los alumnos. Los examinadores jefe y los examinadores principales deben identificar con claridad estas mejoras y ofrecer apoyo al IB para abordarlas.

Los modelos de evaluación se renuevan como parte del proceso de revisión del currículo correspondiente a cada asignatura. Las asignaturas las revisa cada siete años un grupo de revisión compuesto por profesores, examinadores, personal del IB y asesores externos. El de revisión del currículo es un proceso consultivo en el que se hacen circular propuestas entre los colegios autorizados para impartir el Programa del Diploma (PD), a fin de que hagan comentarios según van desarrollándose dichas propuestas. Las recomendaciones y propuestas de los grupos de revisión del currículo también se someten a la consideración del Comité de revisión del Programa del Diploma. Dicho comité es responsable de la calidad académica global de los cursos que componen el PD, y aprueba programas de estudios y los modelos de evaluación propuestos. El comité está particularmente implicado en:

- El estándar académico y la equiparación de los diferentes cursos
- La reducción al mínimo del solapamiento de contenidos y objetivos entre asignaturas, y el fomento de cursos que se complementen unos a otros
- El control de la carga de la evaluación global que han de soportar alumnos, profesores y el IB, para garantizar que puedan sobrellevarla
- La eliminación de la duplicación innecesaria de evaluaciones

En el contexto de la evaluación, tanto los grupos de revisión del currículo como el Comité de revisión del Programa del Diploma toman como referencia la política de evaluación del PD, que define los parámetros dentro de los cuales deben desarrollarse los modelos de evaluación. Así, se deben indicar los problemas que pueda haber con las evaluaciones actuales y se deben tener en cuenta las mejoras que se puedan realizar en los niveles actuales de validez, fiabilidad y justicia.