

Evaluación y Educación en Ciencias Basada en la Indagación:

**Aspectos de la Política y la
Práctica**

Wynne Harlen

Comité Editorial:

**Derek Bell, Jens Dolin, Pierre Léna, Shelley Peers,
Xavier Person, Patricia Rowell y Edith Saltiel**

Traducción: Rosa Devés y Pilar Reyes

Global Network of Science Academies (IAP) Science Education
Programme

Evaluación y Educación en Ciencias Basada en la Indagación:

Aspectos de la Política y la Práctica

Wynne Harlen

Comité Editorial:

**Derek Bell, Jens Dolin, Pierre Léna, Shelley Peers,
Xavier Person, Patricia Rowell y Edith Saltiel**

Traducción: Rosa Devés y Pilar Reyes

Publicado por Global Network of Science Academies (IAP) Science Education Programme (SEP)

www.interacademies.net/activities/projects/12250.aspx

TWAS-Strada Costiera, 11-34151, Trieste, Italia

ISBN: 978-1-291-49836-3

© Wynne Harlen 2013

Las copias y traducciones se pueden hacer sin pago o permiso previo, con el debido reconocimiento.

Evaluación y Educación en Ciencias Basada en la Indagación

Aspectos de la Política y la Práctica

Contenidos

Introducción	1
Contexto del libro	1
Principales puntos que emergen de la conferencia	2
El foco de este libro	5
1 Clarificación de los términos	6
La evaluación y sus distintos propósitos	6
Las pruebas y otros métodos de evaluación del aprendizaje	7
Validez, fiabilidad, recursos y gestión	8
2 Educación en ciencias basada en la indagación: fundamentos y objetivos	12
La indagación en la educación en ciencias	12
Fundamentos	14
Objetivos	15
3 Propósitos y usos de la evaluación del aprendizaje	18
Propósitos y usos	18
Evaluación formativa	19
Evaluación sumativa	23
La relación entre la evaluación formativa y sumativa	26
4 Evaluación del aprendizaje, pedagogía y currículum	29
Impactos de la evaluación	29
Rendición de cuentas	33
Evaluación y teorías del aprendizaje	35
El rol de la evaluación en la comprensión de los principales problemas globales	38
5 Implementación de la evaluación formativa de ECBI	40
Evaluación formativa de ECBI	40
Diálogo en el aula	41
Las preguntas de los profesores	45
Retroalimentación a los estudiantes	48
Retroalimentación hacia la enseñanza	51
Autoevaluación y evaluación entre pares de los estudiantes	51

6	Implementación de la evaluación sumativa de ECBI	56
	Evaluación de la comprensión y las habilidades científicas	56
	Métodos de evaluación sumativa de los objetivos de ECBI	58
	El potencial y las limitaciones de las pruebas para evaluar los objetivos de ECBI	59
	Alternativas a las pruebas	70
	Implementación de la evaluación sumativa hecha por profesores	71
	Mejorando la confiabilidad de la evaluación hecha por profesores	72
	Ejemplos de evaluación hecha por profesores	73
	Informando la evaluación sumativa: pros y contras de utilizar niveles	80
	Ventajas y desventajas de las pruebas y de la evaluación hechas por profesores	82
7	Cambiando las prácticas de evaluación	84
	Aproximaciones hacia el cambio de las prácticas en educación	84
	Algunos ejemplos de cambio en las prácticas de evaluación	86
	Implicancias para la evaluación de ECBI	93
	Bibliografía	101
	Tablas	
1	Estándares para la práctica de la evaluación en el aula	97
2	Estándares para ser utilizados por el equipo de gestión de la escuela	98
3	Estándares para el uso en la Inspección Nacional y Dispositivos de Asesorías	99
4	Estándares para ser usados en la formulación de política pública	100
	Recuadros	
1	Países representados en la Conferencia de Helsinki	2
2	El compromiso entre la validez y la fiabilidad	10
3	Actividades de los estudiantes: aprendiendo a través de la indagación	17
4	Actividades de los estudiantes: aprendiendo a través de transmisión	17
5	Prácticas claves de la evaluación formativa	21
6	Prácticas claves de la evaluación sumativa	25
7	Revisión de la investigación sobre el impacto de la evaluación a gran escala	32
8	Las escuelas haciendo su propio relato	35
9	Un modelo de aprendizaje de las ciencias a través de la indagación	37
10	Explicando el efecto de la presión del aire	42
11	Conversación dialógica sobre los huevos	44
12	Preguntas para descubrir y estimular el desarrollo de ideas	46
13	Preguntas para descubrir y estimular el desarrollo de habilidades de indagación	47
14	Preguntas para estimular la colaboración, compartir ideas, reflexión y evaluación	47
15	Comunicando a estudiantes de enseñanza básica criterios para evaluar su trabajo	53
16	Estudiantes secundarios utilizando autoevaluación para mejorar su trabajo	54
17	Desigualdad de oportunidades en las pruebas	66
18	Fuentes de ítems y tareas para la evaluación en ciencias	67
19	Ejemplos de criterios de evaluación relacionados a la indagación científica	75
20	Criterios para evaluar portafolios (<i>Queensland Senior Certificate</i>)	77
21	Principios de las prácticas de evaluación del aprendizaje	95

Introducción

En los últimos años se ha producido una rápida expansión del interés por la educación en ciencias basada en la indagación (ECBI). Las prácticas de aula y de laboratorio y los materiales que estimulan a los estudiantes a tomar parte activa en la comprensión de los eventos y fenómenos que ocurren en el mundo, se están promoviendo y desarrollando a través de proyectos piloto en países de todo el mundo. Esta aceptación de la educación basada en la indagación, es una manifestación del reconocimiento de su potencial para facilitar que los estudiantes desarrollen comprensión, competencias, actitudes e intereses que son requeridos por todos para vivir en sociedades cada vez más dependientes de las aplicaciones de la ciencia. La indagación conduce al conocimiento de los objetos o fenómenos particulares que se investigan, pero aún más importante, ayuda a construir conceptos generales que tienen poder explicativo amplio, lo que permite entender nuevos objetos o eventos. También genera reflexión sobre los procesos de pensamiento y estrategias de aprendizaje que son necesarios para el aprendizaje continuo a lo largo de la vida. Existen, sin embargo, muchos desafíos en la implementación de ECBI. Uno de los centrales es la evaluación del aprendizaje de los estudiantes, ya que esto tiene una fuerte influencia sobre lo que se enseña y cómo se enseña.

Contexto del libro

La decisión de escribir este libro fue estimulada por una conferencia internacional celebrada en Helsinki, entre el 30 de mayo y el 1 de junio de 2012. La conferencia – que fue planificada en conjunto por la Red Global de Academias de Ciencias (IAP, Global Network of Science Academies), ALLEA (All European Academies), la Academia de Ciencias de Finlandia y el Centro “Letters and Finland’s Science Education” (LUMA) – se tituló Desarrollando la Educación en Ciencias Basada en la Indagación: Nuevos Temas. Los temas concernían el rol de la evaluación en ECBI y la relación de ECBI con la industria. La mayor parte de la conferencia estuvo dedicada a cuestiones relacionadas con la evaluación, que es el tema central de este libro.

El Programa de Educación en Ciencias (SEP, sigla en inglés) de la IAP, bajo el liderazgo de su fundador, el profesor Jorge Allende, ha celebrado importantes conferencias internacionales sobre diversos aspectos de ECBI en 2005, 2008 y 2010, intercaladas con las reuniones de grupos de trabajo para planificar e informar sobre las actividades realizadas. El programa también ha patrocinado actividades a través de cuatro Redes Regionales de las Academias de todo el mundo. Tras el retiro del profesor Allende, el SEP ha sido dirigida por el profesor Pierre Lena desde 2011.

A lo largo del trabajo del IAP-SEP se ha reconocido que el rol de la evaluación de los estudiantes se vuelve cada vez más importante para la comprensión y aplicación de ECBI. La evaluación es un aspecto clave en la planificación estratégica para el cambio educativo. A raíz de la expansión de los proyectos piloto se ha producido una demanda de información sobre la eficacia de ECBI para justificar los recursos necesarios para su implementación. Para ello se requiere de buenas estimaciones sobre los resultados de ECBI. Pero suministrar información sobre los logros de los estudiantes es sólo uno de los roles de la evaluación; su rol en cuanto a apoyar el aprendizaje y desarrollar una comprensión más profunda en relación con los objetivos de la enseñanza de las ciencias ha ganado un apoyo considerable en los últimos años en todo el mundo.

Recuadro 1: Países representados en la conferencia de Helsinki

Argentina
Australia
Austria
Brasil
Camerún
Canadá
China
Colombia
Costa Rica
Dinamarca
Estados Unidos de América
Eslovenia
Estonia
Finlandia
Francia
Alemania
Ghana
Haití
Hungría
India
Irán
Israel
Italia
Mauricio
México
Kenia
Kosovo
Malasia
Montenegro
Mozambique
Holanda
Nueva Zelanda
Nigeria
Noruega
Pakistán
Reino Unido
Senegal
Serbia
Sud Africa
Sudán
Suecia
Suiza
Tanzania
Túnez
Uganda
Ucrania
Venezuela
Vietnam
Zambia

Además, el reconocimiento de que lo que se evalúa influye el contenido del currículo y la pedagogía, apunta a que la evaluación también puede retrasar el cambio en la educación en ciencias, cuando los métodos de evaluación establecidos y los contenidos no reflejan las metas de ECBI.

Hacer frente a los problemas que inhiben la implementación del ECBI se hace más importante cuando se considera el valor que representa el aprendizaje basado en la indagación para el desarrollo de aquellas habilidades que son requeridas por la fuerza de trabajo actual y futura. Por estas razones, el Consejo Mundial de la IAP SEP, en su reunión celebrada en París en abril de 2011 acordó que la evaluación de los estudiantes sería un foco importante de su conferencia de 2012 y condujo a la decisión de producir este libro.

Principales puntos que emergen de la conferencia

Las presentaciones de la conferencia de Helsinki, y en particular las discusiones entre los 93 participantes de 50 países (véase el Recuadro 1), dieron cuenta de una serie de preocupaciones relativas a la evaluación de los alumnos, brevemente esbozadas aquí en seis secciones:

- La necesidad de clarificar los términos
- Comprendiendo las interacciones entre la evaluación, la pedagogía, los contenidos curriculares, y los marcos de política educativa
- El rol de las pruebas, las tareas y los profesores en la evaluación de ECBI
- La adecuación de la evaluación para distintos propósitos
- Los desafíos de hacer cambios en la evaluación
- La necesidad de más investigación en el área de la evaluación.

La necesidad de clarificar los términos

Si vamos a avanzar en el desarrollo de políticas y prácticas en relación con la evaluación de ECBI es muy importante que exista una comprensión común sobre los términos utilizados. La discusión en la conferencia puso en evidencia cierta confusión sobre el significado de términos claves, tales como evaluación del aprendizaje (*assessment*, en inglés), realización de pruebas (*testing*) y evaluación de programas (*evaluation*), y sobre la diferencia entre la evaluación del aprendizaje formativa y sumativa. Las diferencias de idioma pueden constituirse en un problema en este caso, sobre todo cuando éste tiene una sola palabra para referirse a los distintos tipos de evaluación (como es

el caso del español, ver nota de traducción en página 6). También se puso en evidencia que el significado de ECBI sigue siendo interpretado de manera algo estrecha como si se tratase sólo del desarrollo de habilidades. Esto es comprensible dado que ECBI ha sido introducido en algunos países como un antídoto a la enseñanza basada en libros de texto. Sin embargo, el tener una visión común sobre lo que significa ECBI en la práctica y en qué se diferencia de otros enfoques de la enseñanza y el aprendizaje, no sólo es necesario para su implementación, sino también esencial como plataforma para el desarrollo de la evaluación. Los participantes en la conferencia señalaron que, más allá de los significados de las palabras, hay diferencias culturales que subyacen al discurso sobre ECBI y su evaluación, lo que lleva a diferentes visiones sistémicas que deben ser reconocidas y debatidas.

Comprendiendo las interacciones entre la evaluación, la pedagogía, los contenidos curriculares, y los marcos de política educativa

Hubo consenso general en que aquello que se evalúa influye en la prioridad dada por los profesores a diversos objetivos de aprendizaje y, por lo tanto, es esencial que todos los objetivos importantes se incluyan en lo que se evalúa. Se sugirió que esto se facilitaría si el currículum y la evaluación fuesen contruidos juntos por las mismas agencias, para evitar que los objetivos fundamentales fuesen convertidos en una serie de tareas inconexas. Un punto relacionado, que fue destacado por varios participantes, es que la evaluación debe ser consistente con la teoría del aprendizaje que subyace a ECBI. Hubo una fuerte opinión de que la conexión entre una buena evaluación, la implementación del ECBI y el desarrollo de ideas científicas claves debe ser explicitada. Se requiere una descripción de lo que se entiende por "calidad" del aprendizaje y sobre cómo esto se entiende en la medida que estudiantes de distintas edades progresan en el aprendizaje "de la ciencia" y "sobre la ciencia". La evaluación debe apoyar, y ser vista como apoyando, el desarrollo de la buena ciudadanía y de los conocimientos y habilidades necesarias para abordar los principales problemas mundiales. Las políticas educativas basadas en la ambición de alcanzar altos niveles de desempeño en las pruebas internacionales tienen que ser reconciliadas con la preocupación por una educación de calidad para todos.

El rol de las pruebas, las tareas y los profesores en la evaluación de ECBI

Los participantes en la conferencia reconocieron que la naturaleza de ECBI plantea muchos desafíos conceptuales, logísticos y técnicos para la evaluación de los alumnos. Conceptualmente, los procesos de evaluación deben ser capaces de comunicar quien ha aprendido bien ciencias. Esto implica claridad sobre los objetivos de la educación científica y sobre la forma en que ECBI contribuye a alcanzar estos objetivos. Técnicamente, los retos son garantizar la fiabilidad, sin comprometer la validez. Los objetivos que buscan desarrollar aprendices seguros, autónomos y colaborativos son difíciles de evaluar en forma directa y es necesario buscar medidas indirectas que reflejen estas cualidades. Se reconoció que es casi imposible obtener, a través de pruebas de longitud razonable, toda la rica información necesaria para evaluar las metas ECBI. Por otra parte, la producción y la administración de pruebas implica un alto costo en tiempo, así como de otros recursos. Los participantes afirmaron que los docentes deberían estar involucrados en conducir la evaluación en algunas o todas las etapas de la escolaridad. Se sugirió que la retroalimentación sobre los resultados de las evaluaciones en general y la evaluación de desempeño de los profesores, podrían tener un papel en el fortalecimiento de las capacidades de evaluación de los docentes. También debe reconocerse el impacto de la cultura escolar. La adopción de nuevos enfoques para la enseñanza y la evaluación puede verse limitada por las medidas de rendición de cuentas que tienen una visión estrecha sobre la enseñanza y el aprendizaje de las ciencias.

Adecuación de la evaluación a los distintos propósitos

Para que la evaluación pueda utilizarse para facilitar el aprendizaje es necesario que los profesores incorporen estrategias de evaluación formativa como parte de su pedagogía en lugar de sumar una serie de eventos de evaluación mini-sumativa. En el caso de la evaluación sumativa, las pruebas se utilizan comúnmente para el control del desempeño al final de ciertos tópicos o cursos y para la elaboración de informes sobre los progresos realizados a intervalos regulares. Hubo amplio acuerdo en que la mayoría de las pruebas de ciencias utilizadas por los docentes no reflejan las metas de ECBI. Dado que las evaluaciones tienen una fuerte influencia sobre lo que se enseña, es importante considerar alternativas a las formas actuales de pruebas con el fin de obtener una información más confiable sobre el aprendizaje que resulta de ECBI. Algunos participantes en la conferencia expresaron la opinión de que el uso de resultados de las pruebas para los efectos de la rendición de cuentas inhibe el aprendizaje. Se observó que los programas de pruebas a gran escala como TIMSS y PISA utilizan métodos que no sirven como inferencias válidas sobre el desempeño en ECBI a nivel del sistema. Sin embargo, se reconoce que los resultados de estas pruebas sí tienen un rol, pero deben considerarse como un indicador más de desempeño del sistema junto a la información de otros aspectos.

Los desafíos de realizar cambios en la política y la práctica de la evaluación

De acuerdo a la percepción de los participantes de la conferencia, uno de los mayores desafíos respecto a los cambios en evaluación se refiere a la participación de los profesores en ésta. Se considera que los profesores necesitan ayuda para desarrollar una "alfabetización en evaluación" y que varios de los objetivos de ECBI serán mejor evaluados por los profesores. Se reconoce que las oportunidades de aula que tienen que ser proporcionadas a los estudiantes para aprender a través de ECBI son también oportunidades para que los profesores evalúen y registren el desarrollo de la comprensión y de las habilidades. Los profesores también necesitan herramientas para apoyar su evaluación formativa y sumativa. En muchos países existen desafíos en relación a la participación de las niñas en la ciencia, pero no existen evidencias de diferencias de género en la competencia en ECBI. Es importante influir en la formación inicial de profesores de manera que los docentes que accedan a la profesión estén conscientes de ECBI y de cómo debe ser incluida en su práctica de evaluación. Los delegados de la conferencia también aconsejaron que los políticos, administradores, padres y el público en general necesitan ser educados acerca de los significados, propósitos, fortalezas y limitaciones de la evaluación.

La necesidad de más investigación relacionada a la evaluación

La investigación reportada en las presentaciones de la conferencia proporcionó mucha evidencia sobre la necesidad de repensar ciertos aspectos de cómo los estudiantes deben ser evaluados para darles la oportunidad de demostrar lo que saben y pueden hacer. La confianza en la precisión de ciertos resultados de la evaluación es a menudo cuestionada por investigaciones que muestran que la forma en que los estudiantes abordan una pregunta o tarea y, los procesos de pensamiento que utilizan para enfrentarlas, no son lo que se suponían cuando se interpretaron los resultados. Se requiere más de estos estudios para mejorar la validez de los métodos y procesos de evaluación, y para disuadir supuestos poco realistas sobre la precisión. Se sabe que incluso los pequeños cambios en el formato y en la elaboración de preguntas escritas pueden influir en el desempeño de los estudiantes, pero menos se sabe sobre el impacto diferencial en los estudiantes de distinto sexo, experiencia y contexto. Varias otras áreas donde se necesita investigación también fueron mencionadas. Estas incluyen: el impacto de los métodos alternativos a las pruebas, la combinación de diferentes tipos de evidencias que proporcionan la mejor imagen de ciertos tipos de desempeño de los estudiantes, cómo puede incrementarse la confianza en la evaluación realizada por los profesores, cómo capacitar a los docentes, tanto en el uso de la evaluación formativa como sumativa, cómo

combinar la evaluación formativa y la evaluación sumativa en los sistemas nacionales y distritales de evaluación.

El foco de este libro

En este libro, tal como en la conferencia, hemos distinguido entre dos conjuntos de problemas relativos a:

- Los procesos y los usos de la *evaluación del aprendizaje de los estudiantes* y
- El *valor* de la educación en ciencias basada en la indagación y la *evaluación de su efectividad*

El primero de estos se refiere a los temas presentados antes que se derivan de reconocer el rol que juega la evaluación de los estudiantes como una parte integral de sus experiencias de aprendizaje. La evaluación se consideró alguna vez como algo que ocurre después del aprendizaje y como independiente del proceso de aprendizaje. Este punto de vista ya no es sostenible, la evaluación se reconoce ahora como una parte central de la educación, con un papel demostrado tanto en cuanto a apoyar el aprendizaje como en informar sobre él. Se reconoce que el cómo se utilicen los resultados de la evaluación de los estudiantes tiene una importante influencia que puede ser positiva o negativa, sobre los contenidos y métodos de enseñanza. Así, la naturaleza de la evaluación, y en particular la medida en que permite a los estudiantes demostrar lo que saben y pueden hacer, en relación con los objetivos de aprendizaje previstos, son factores claves en la educación de los estudiantes. En el contexto de ECBI es un motivo de preocupación que la mayoría de las herramientas de evaluación y los procedimientos actuales no están a la altura de lo que se necesita para dar cuenta del logro de los estudiantes en cuanto a sus objetivos.

En relación con el segundo aspecto, el valor de ECBI no es un asunto que pueda decidirse por la evidencia empírica, sino que es un juicio de valor en cuanto a que las competencias, la comprensión, los intereses y las actitudes que incluyen sus objetivos valen la pena y son necesarios en una educación moderna. Aquello que la evaluación de programas *puede* mostrar es el grado en que los estudiantes logran estos objetivos a través de experiencias diseñadas para implementar ECBI. La evaluación de los estudiantes tiene un rol en la evaluación de los programas, pero hay muchos otros factores involucrados. En particular, lo que los estudiantes logren será informativo en relación con los resultados de ECBI, sólo si existe evidencia de que los estudiantes están de verdad aprendiendo a través de la indagación y cuando los datos aportados por la evaluación permitan realizar inferencias sobre la comprensión de la ciencia y las habilidades de indagación, que son los objetivos de ECBI. La naturaleza de la evaluación del aprendizaje de los estudiantes en ciencias es uno de los principales factores que frenan la aplicación de los programas ECBI.

Es indudable que existe una necesidad de conocer el impacto de la educación basada en la indagación sobre los logros de los estudiantes, pero esto no puede hacerse sin herramientas válidas. Así, al focalizar aquí en la primera serie de cuestiones mencionadas anteriormente – la evaluación del aprendizaje de los estudiantes – no sólo se ofrece la promesa de proveer algunas de las herramientas necesarias para la evaluación de programas, sino, sobre todo, considerar la evaluación como una parte integral del aprendizaje de los estudiantes a través de la indagación.

Capítulo 1

Clarificación de los términos

Muchos de los términos utilizados en la discusión sobre evaluación tienen tanto un uso técnico como un uso común, como es el caso de muchos de los términos usados en ciencias. La falta de claridad sobre su significado, así como de consistencia en su uso, impiden la buena comunicación y la comprensión de las diferencias y las relaciones entre los conceptos. Si bien no todos estarán de acuerdo en la definición de los términos utilizados en la evaluación - y entendiendo que existirán problemas específicos con la traducción del inglés a otros idiomas - es importante, por lo menos, explicitar el significado que se da a las palabras utilizadas en este libro. En particular, a términos como "evaluación del aprendizaje" (*assessment*, en inglés), "realización de pruebas" (*testing*), "evaluación de programas" (*evaluation*) y las palabras asociadas, tales como estándares, criterios, validez y fiabilidad. Los significados de indagación se discuten en el Capítulo 2 y en el Capítulo 3 donde nos referimos a otros términos relacionados con los fines, funciones y usos de la evaluación.

(Nota de traducción: considerando que en español el término "evaluación" se utiliza para todos los casos, nos hemos preocupado de distinguir los distintos ámbitos agregando especificaciones cuando sea necesario, tales como: evaluación del aprendizaje (formativa o sumativa), evaluación de programas, evaluación de escuelas, evaluación de políticas, etc. Puesto que el foco del libro es la evaluación del/ para el aprendizaje, frecuentemente usamos el término evaluación a secas, para referirnos a estos procesos, cuidando que su significado preciso quede claro en el contexto en que se usan. Lo mismo vale para el título del libro).

La evaluación y sus distintos propósitos

La OECD, en sus revisiones sobre la evaluación en sus países miembros, hace una distinción clara y útil entre la evaluación del aprendizaje de los estudiantes, la evaluación del desempeño de los profesores, la evaluación de escuelas y la evaluación del sistema, al mismo tiempo que reconoce que estos componentes del sistema tienen que trabajar juntos en las políticas que tienen como objetivo mejorar los logros de aprendizaje:

El término "evaluación del aprendizaje" (*assessment*) se utiliza para referirse a los juicios sobre el desempeño individual del estudiante y el logro de objetivos de aprendizaje. Este concepto cubre la evaluación basada en el aula, así como los tests externos y pruebas de gran escala. El término "evaluación de desempeño" (*appraisal*) se utiliza para referirse a los juicios sobre el desempeño de los profesionales a nivel de escuela, e.g. profesores y directores. Finalmente, el término "evaluación" (*evaluation*) se utiliza para referirse a los juicios sobre la efectividad de las escuelas, los sistemas escolares y las políticas¹.

Tanto la evaluación del aprendizaje como la evaluación de programas describen un proceso de generación e interpretación de evidencias para algún propósito. Ambos involucran tomar decisiones acerca de qué evidencia utilizar, generar y recoger evidencia en forma sistemática y planificada, interpretar la evidencia para producir un juicio, y comunicar y utilizar ese juicio. Vale la pena señalar

1 Nusche, D. et al (2012) *OECD Reviews of Evaluation and Assessment in Education: New Zealand 2011*. OECD Publishing, p24

que la evidencia, de cualquier tipo, siempre es sólo una indicación o una muestra de un rango más amplio de evidencias que podrían ser utilizadas.

En este libro, de acuerdo a la convención de la OECD, el término "evaluación del aprendizaje" se usa para referirse al proceso de recoger y usar la evidencia sobre los logros de aprendizaje - por lo general el aprendizaje de los estudiantes - pero también puede referirse al aprendizaje de otros, tales como profesores.

El término evaluación (*evaluation*) se utiliza (en la versión original en inglés, nota de traducción) en relación a la generación y uso de la evidencia acerca de los sistemas, materiales, procedimientos y procesos. La evaluación de las escuelas, sistemas y enfoques de la enseñanza pueden hacer uso de la evidencia del aprendizaje de los estudiantes, pero el juicio es sobre el valor o el éxito de otras cosas, tales como las políticas escolares y los programas, más que el aprendizaje de los estudiantes, aunque esto pueda ser parte de la evidencia utilizada en la evaluación.

Las pruebas y otros métodos de evaluación del aprendizaje

Aunque los términos evaluación del aprendizaje y examinación (o realización de pruebas) se utilizan a veces indistintamente existe una diferencia importante entre ellos. Las pruebas pueden ser consideradas como un método de recolección de datos para la evaluación del aprendizaje, sin embargo, la evaluación del aprendizaje es un concepto más amplio que abarca otros métodos de recopilación e interpretación de datos, además de la realización de pruebas.

Una mirada más cercana a lo que involucra la evaluación del aprendizaje ayuda a clarificar esta relación y a identificar otros aspectos de la evaluación que incluyen términos como "estándares" y "criterios".

Toda evaluación de los logros de los estudiantes involucra la generación, interpretación, comunicación y uso de datos para algún propósito. Sólo en esta simple declaración hay espacio para una enorme gama de diferentes tipos de actividades, pero cada una de ellas implicará: a) estudiantes involucrados en alguna actividad, b) la recolección de datos de esa actividad por parte de algún agente, c) el juicio sobre la datos, comparándolos con algún estándar y d) algunos medios para describir y comunicar el juicio. Cada uno de los componentes de la evaluación del aprendizaje puede tomar diversas formas.

- a) Actividades en las cuales pueden estar involucrados los estudiantes:
 - su trabajo regular
 - pruebas escritas o prácticas creadas por el profesor o profesora con el propósito de evaluación
 - pruebas escritas o prácticas creadas externamente
- b) Los datos pueden ser recogidos por:
 - el(la) profesor(a)
 - los(as) estudiantes
 - el(la) profesor(a) y los estudiantes juntos
 - un agente externo (comisión examinadora, autoridad acreditadora, desarrollador de pruebas).
- c) Los datos pueden ser juzgados en relación a:
 - normas, en cuyo caso el estándar de comparación es el desempeño de otros estudiantes (referenciado a la norma)
 - criterios, en cuyo caso el estándar de comparación es la descripción de aspectos del desempeño (referenciado a criterios)
 - desempeños anteriores de los estudiantes, en cuyo caso el desempeño individual se juzga en relación a los desempeños anteriores (u otros) de los estudiantes (referenciado al estudiante o ipsativo).

d) El juicio debe ser comunicado como:

- un comentario escrito u oral por parte del profesor o profesora
- una nota o puntaje o porcentaje
- un perfil de logro
- un nivel o grado
- un rango o percentil

Es posible crear diferentes herramientas y procedimientos de evaluación del aprendizaje a través de diferentes combinaciones de estas diversas formas de recoger, juzgar y comunicar datos. Por ejemplo, una prueba estandarizada comprende tareas creadas por una agencia externa que ha puesto a prueba el test durante su desarrollo con una amplia muestra de la población correspondiente, de manera que el puntaje individual puede expresarse en términos de comparación con la "norma" para esa población. El resultado indicará si el desempeño de un estudiante está por encima o por debajo de la media, pero no es lo que él o ella es capaz de hacer.

Una prueba referenciada a criterios difiere de una prueba referenciada a una norma en cuanto ha sido diseñada para dar información respecto a lo que un estudiante puede hacer en relación a logros especificados. Los ítems serán elegidos por su relevancia para el currículum de manera que los resultados puedan ser utilizados para establecer, no cómo un estudiante se compara con otros, sino cómo su desempeño se compara con el desempeño previsto. Al mismo tiempo, el nivel esperado de logro se establece en referencia a lo que se puede esperar de la población a la que está destinada la prueba. De esta manera, existe un elemento normativo en la decisión sobre los criterios contra los cuales se juzga el desempeño². Cuando se usan las pruebas o tests, los datos se restringen a los ítems de la prueba, mientras que si la evaluación la llevan a cabo los profesores existe el potencial de utilizar toda la gama de actividades de aprendizaje para emitir juicios usando los criterios de desempeño esperado en relación a los objetivos de la lección.

Validez, fiabilidad, recursos y gestión

Para decidir la mejor manera de llevar a cabo la evaluación del aprendizaje en un caso particular, es necesario tener en cuenta las propiedades de las posibles herramientas en relación con los propósitos y usos que se darán a los resultados de la evaluación. Una propiedad deseable es que cualquier evaluación deberá ser válida para sus propósitos; esto es deberá evaluar lo que se pretende evaluar. Otra es que debe proporcionar datos fiables, o dignos de confianza. Pero también hay otras cuestiones a tener en cuenta, en particular, si se tiene a la vista la interdependencia de los diversos componentes del sistema, el impacto sobre otras prácticas de evaluación del aprendizaje, sobre el currículum y sobre la pedagogía. Además deben considerarse, el uso de los recursos - la evaluación puede ser costosa, tanto en términos de recursos monetarios como del tiempo de los estudiantes y profesores - y los aspectos relativos a la gestión.

Validez

Se suele definir la validez de una evaluación del aprendizaje en términos de qué tan bien aquello que se evalúa se corresponde con el comportamiento o los resultados de aprendizaje que se pretende evaluar. Se han propuesto varios tipos de validez en función del tipo de información que se utiliza para juzgar la validez. Por ejemplo, la *validez de contenido* se refiere a qué tan adecuadamente la evaluación cubre el dominio de la materia que se enseña y suele basarse en el juicio de los expertos en el tema. Sin embargo, la cobertura de contenido no es suficiente para distinguir un test u otra evaluación de la ciencia que ha sido aprendida a través de la indagación, de la evaluación de la ciencia aprendida de otra manera. La *validez de constructo* es un concepto más amplio, que refleja

2 Black, P (1998) Testing: Friend or Foe? London: Falmer Press.

el rango completo de logros de aprendizaje en un particular dominio del conocimiento. El requisito importante es que la evaluación muestree todos los aspectos de los logros de los estudiantes que sean relevantes para el propósito específico de la evaluación. Debe tenerse en cuenta, que el incluir aspectos irrelevantes amenaza la validez, de la misma forma que omitir aspectos relevantes.

Sin embargo, la visión de la validez como una propiedad de un método o instrumento de evaluación del aprendizaje, sin considerar las circunstancias en que se utiliza y los usos que se hacen de los resultados, ha sido ampliamente cuestionado. Newton³ ha señalado el error en estos dos supuestos. En el caso de las circunstancias en que se utiliza la evaluación, la exactitud de los resultados como una medida del constructo dependerá de cómo la evaluación se administra, así como de qué es lo que contiene. En el caso de la utilización de los resultados, puede que se hagan afirmaciones respecto a lo que se intentaba evaluar, cuando en realidad otros factores fueron más influyentes en el resultado (como cuando una prueba de matemáticas tiene una alta demanda de lectura, de manera que es incierto si es la lectura o la habilidad matemática lo que más influye en los resultados.)

La noción de validez que tiene en cuenta no sólo lo bien que la evaluación muestrea el constructo que se intenta evaluar, sino qué es lo que se afirma en base de los resultados, es aquella que se relaciona a las inferencias que se deducen de los resultados. Esto fue expresado formalmente en una definición de validez de Messick que ha sido ampliamente citada:

La validez es un juicio evaluativo integral del grado en que la evidencia empírica y los fundamentos teóricos apoyan la adecuación e idoneidad de las inferencias y acciones basadas en los resultados de pruebas u otros modos de evaluación⁴.

Una consecuencia de adoptar este punto de vista de la validez es reconocer que la validez puede variar debido a factores que afectan el desempeño, tales como las condiciones en que se examina, y que se describen generalmente en términos de la fiabilidad de la evaluación. Volveremos a este punto después de considerar el significado de fiabilidad.

Fiabilidad

La fiabilidad de la evaluación se refiere al grado en que los resultados pueden considerarse de una consistencia o precisión aceptable para un uso particular. Este puede no ser el caso si, por ejemplo, los resultados son influenciados por quien realiza la evaluación o si dependen de la ocasión particular o las circunstancias de un momento determinado. Así, la fiabilidad se define a menudo como y se mide por, el grado en que la evaluación, si se repitiera, daría el mismo resultado.

La fiabilidad tiene sentido principalmente en el caso de la evaluación sumativa y en particular en el caso de las pruebas. Cuando la evaluación se usa formativamente (ver Capítulo 3), involucra sólo a los estudiantes y a sus profesores, y la noción de realizar un juicio repetible y de tratar a todos los estudiantes de la misma manera no es relevante. No existe, en este caso, un juicio de grado o nivel, sólo el juicio de cómo ayudar al estudiante a tomar los próximos pasos en el aprendizaje, por lo que la fiabilidad en este sentido formal no es un tema. Para la evaluación formativa lo que es importante es "la calidad de la información que se reúne y que se ofrece como retroalimentación".⁵

3 Newton, P. Validity, purpose and the recycling of results. In (ed) J. Gardner *Assessment and Learning*. 2nd edn. London: Sage

4 Messick, S.(1989) Validity, in (ed) R. Linn *Educational Measurement* (3rd edn) American Council on Education, Washington: Macmillan, pp 13-103, p13.

5 Stobart, G. (2012) Validity in formative assessment In (ed) J. Gardner *Assessment and Learning*. 2nd edn. London: Sage p234.

Recuadro 2: El compromiso entre la validez y la fiabilidad

En cualquier evaluación existe un límite respecto al grado de optimización de la fiabilidad y la validez. Esto aplica a cualquier forma de evaluación, pero se identifica más claramente en relación al uso de los tests. Al seleccionar los ítems para un test, y con el fin de aumentar la fiabilidad, es inevitable que la selección dé preferencia a aquellos ítems que puedan ser calificados en forma consistente o cuya calificación puede ser automatizada. Esto favorece a los ítems que evalúan conocimiento factual y que usan un formato cerrado, en oposición a los ítems que requieren la aplicación de conocimiento y el uso de tareas más abiertas. La consecuente limitación en lo que se cubre en una prueba afecta a su validez.

Los intentos de aumentar la validez ampliando el rango de ítems, por ejemplo, incluyendo más preguntas abiertas en las cuales se requiere de mayor juicio para evaluar, llevan a una reducción de la fiabilidad. Por ello, existe un compromiso entre la fiabilidad y la validez; aumentando la fiabilidad disminuye la validez y vice-versa.

Sin embargo, una alta confiabilidad es necesaria cuando los resultados son utilizados por otros y cuando los estudiantes están siendo comparados o seleccionados. Así, la discusión aquí se refiere a los tests, en los que la fiabilidad es importante y se asume que todos los estudiantes experimentan las mismas condiciones y tienen las mismas oportunidades. Por cierto que nunca es el caso que las mismas condiciones signifiquen las mismas oportunidades para estudiantes diferentes, o que un estudiante individual reaccione de la misma manera a una prueba en ocasiones diferentes. Las respuestas individuales de los estudiantes a las mismas condiciones variarán de un día a otro y, aún más importante, la selección particular de ítems de una prueba puede resultar más difícil para algunos estudiantes que otros de la misma capacidad. Todos los ítems de una prueba presentan preguntas en un cierto contexto y existe evidencia empírica de que los estudiantes que tienen un buen desempeño en un ítem no lo harán bien en otro que somete a prueba los mismos conceptos o habilidades, pero en un contexto diferente. Para cualquier prueba existe un gran número de ítems posibles y sólo una pequeña muestra de ellos puede incluirse en un test de longitud razonable. Una selección diferente produciría un resultado

diferente, dando lugar a lo que se describe como el "error de muestreo".

El error de muestreo puede ser mucho mayor de lo que generalmente se cree. Por ejemplo, Wiliam⁶ ha estimado que en las pruebas nacionales en Inglaterra alrededor del 40% de los estudiantes son asignados al nivel (o grado) equivocado, a pesar de que cada uno de éstos abarca aproximadamente dos años. Una forma de reducir esta fuente de error sería incrementar el número de contextos que se incluyen para cada competencia evaluada y, por lo tanto, el número de ítems utilizados. Pero la longitud de una prueba no se puede aumentar mucho sin incurrir en otras formas de error (fatiga del estudiante, por ejemplo) de manera que más ítems por habilidad significaría incluir menos habilidades o conceptos, reduciendo así el rango de lo que se evalúa y la validez de la prueba. Este es un ejemplo de la interacción entre la fiabilidad y validez identificada en el Recuadro 2. La consecuencia es mayor en relación a las pruebas de testeo individual en las que todos deben recibir los mismos ítems. El efecto es mucho menor en relación a los estudios poblacionales en los que el error de muestreo puede reducirse usando un gran número de ítems distribuidos en muestras equivalentes de estudiantes que

6 Wiliam, D. (2001) Reliability, validity and all that jazz, *Education* 3-13, 29 (3): 17-21.

abordan distintos grupos de ítems (véase el Capítulo 6).

Recursos y gestión

Los recursos necesarios para realizar una evaluación deben ser acordes con el valor de la información para los usuarios de los datos. Los recursos pueden ser el tiempo de los profesores, la experiencia y el costo tanto para la escuela, como para los organismos externos que participan en la evaluación. En general debe existir un compromiso, en particular cuando se requiere un alto grado de precisión. Existe un límite en cuanto al tiempo y la experiencia que puede utilizarse en el desarrollo y la operación de, por ejemplo, una prueba externa o test altamente fiable. La triple revisión de todos los exámenes claramente se generaría una mayor confianza en los resultados, observadores visitando a todos los candidatos aumentaría la gama de resultados que pueden ser evaluados externamente, la capacitación a todos los profesores para actuar como asesores expertos tendría grandes ventajas - pero no todos éstos son realistas en práctica. Equilibrar costos y beneficios plantea cuestiones de valores, así como de posibilidades técnicas.

El costo de la evaluación formativa es insignificante una vez que se ha incorporado a la práctica. El proceso de introducción puede ser considerable en términos del tiempo de los profesores para el desarrollo profesional. Una buena evaluación formativa, como se explica en el Capítulo 5, requiere no sólo del dominio de ciertas estrategias de aula, sino también conocimiento sobre las rutas de progresión en aspectos del aprendizaje, y ejemplos de maestros y estudiantes utilizando evidencias para identificar los próximos pasos en el aprendizaje. Estos costos, sin embargo, son parte integral de los esfuerzos para mejorar el aprendizaje.

La evaluación sumativa requiere recursos, tanto en términos de los tiempos de los profesores, como de los alumnos. Cuando se usan pruebas desarrolladas por agencias externas a la escuela o por editoriales comerciales, hay un costo considerable. Incluso si las pruebas y los exámenes nacionales son gratuitos para las escuelas, el costo debe ser soportado por el sistema y puede ser sorprendentemente grande. Si los costos directos de producir, distribuir y calificar las pruebas se agregan al tiempo empleado para prepararse para y dar las pruebas y exámenes externos, el total puede ascender a una proporción significativa del presupuesto de educación.⁷ Sin duda muestra la importancia de considerar el equilibrio entre costos y beneficios al decidir los métodos que se utilizarán para la evaluación sumativa.

7 Harlen, W. (2007) *Assessment of Learning*. London: Sage p 61/2

Capítulo 2

Educación en ciencias basada en la indagación: fundamentos y objetivos

Las decisiones acerca de las maneras más efectivas de garantizar que la evaluación se utilice para apoyar e informar el aprendizaje de las ciencias a través de la indagación requieren de una clara identificación de los resultados de aprendizaje que se esperan de ECBI. De esta manera se podrán seleccionar o diseñar los enfoques de evaluación que aporten los datos más confiables sobre estos resultados. En este capítulo nos preocupamos de los objetivos de ECBI y las razones que determinan la importancia de estos objetivos. Antes de embarcarnos en esto, sin embargo, es importante tener en cuenta que la indagación no es el único enfoque utilizado en la educación científica. Hay aspectos del aprendizaje de las ciencias, como el conocimiento del vocabulario científico, las convenciones y el uso de los equipos, que se aprenden mejor a través de la instrucción directa. Así, no toda la enseñanza científica y no toda la evaluación deberán estar concernidas con logros específicos del aprendizaje a través de la indagación. Sin embargo, el conocimiento de hechos y procedimientos son medios para el propósito final de desarrollar comprensión a través de la indagación, por lo tanto, el elemento decisivo en la evaluación debe reflejar la comprensión, las habilidades y las competencias que son las metas de ECBI.

La indagación en la educación en ciencias

La indagación es un término que se utiliza tanto en la educación como en la vida cotidiana para referirse a la búsqueda de explicaciones o información a través de preguntas. A veces se equipara con la investigación o la "búsqueda de la verdad". Dentro de la educación, la indagación puede aplicarse en distintos dominios temáticos, como la historia, la geografía, las artes así como en la ciencia, las matemáticas, la tecnología y la ingeniería, cuando se plantean preguntas, se recoge evidencia y se consideran las posibles explicaciones. En cada área emergen diferentes tipos de conocimiento y comprensión. Lo que distingue a la *indagación científica* es que conduce al conocimiento y la comprensión del mundo natural y artificial a través de la interacción directa con el mundo y a través de la generación y recolección de datos para su uso como evidencia en el proceso de someter a prueba las explicaciones de fenómenos y eventos.

La indagación no es en ningún caso un concepto nuevo en educación, al basarse en el reconocimiento del rol activo de los niños en el desarrollo de sus ideas y de su comprensión. Los estudios de Piaget⁸ y los argumentos de Dewey⁹ entre otros, en la primera mitad del siglo 20 llamaron la atención al importante rol de la curiosidad, la imaginación y la avidez de interactuar y preguntar en el aprendizaje de los niños. Más recientemente, el National Research Council de los EE.UU. ha señalado el valor que tiene que los estudiantes se involucren en realizar observaciones, plantear preguntas, usar herramientas para recopilar, analizar e interpretar datos y comunicar los resultados.¹⁰ Del mismo modo, la National Science Foundation de los EE.UU. definió la enseñanza basada en la indagación como aquella que lleva a los estudiantes a desarrollar su comprensión de las

8 Piaget, J (1929) *The Child's Conception of the World*. New York: Harcourt Brace.

9 Dewey, J. (1933) *How we think: A restatement of the relation of reflective thinking to the educative process*. Boston, MA: D.C. Heath

10 National Research Council (NRC) (1996) *National Science Education Standards*. Washington DC: National Academy Press

ideas científicas fundamentales a través de la experiencia directa con los materiales, y la consulta de libros, otros recursos y expertos, y a través de la discusión y debate entre ellos. ¹¹

En el curso de varios proyectos piloto de la última década, el Programa de Educación en Ciencias del IAP ha formulado la siguiente definición de la educación en ciencias basada en la indagación:

ECBI significa que los estudiantes desarrollan progresivamente ideas científicas claves mientras aprenden a investigar y construyen su conocimiento y comprensión del mundo que los rodea. Ellos utilizan habilidades empleadas por los científicos tales como hacer preguntas, recoger datos, razonar y revisar evidencia a la luz de lo que ya se conoce, extraer conclusiones y discutir los resultados. Este proceso de aprendizaje está apoyado por una pedagogía basada en la indagación, donde la pedagogía se entiende no sólo como el acto de enseñar, sino también como las justificaciones que lo sustentan.¹²

Algunos términos en esta definición merecen énfasis y comentario.

Desarrollo progresivo de ideas claves subraya la importancia de identificar algunas ideas generales que nos ayuden a dar sentido a los fenómenos del mundo que nos rodea, y luego asegurarse de que a través de las actividades de aprendizaje de la ciencia, los estudiantes avancen hacia el desarrollo de estas ideas.

Construir su conocimiento y comprensión implica el rol activo de los estudiantes en el aprendizaje, lo que es parte de la evaluación formativa, discutida en el Capítulo 3, y que implica una perspectiva del aprendizaje como algo construido por los aprendices, descrita en el Capítulo 4.

Utilizar las habilidades empleadas por los científicos significa, además de las habilidades indicadas, ser riguroso y honesto en la recopilación y uso de datos suficientes y pertinentes para someter a prueba las hipótesis o responder a las preguntas planteadas. Los científicos comprueban y repiten la recolección de datos, cuando es posible, interpretan y tratan de explicar sus hallazgos. A lo largo de sus investigaciones mantienen un registro cuidadoso, y en la elaboración de conclusiones consultan trabajos relacionados y presentan su trabajo a los demás por escrito o en conferencias, y comparten sus ideas. Algo que es obvio en el caso de los científicos, pero que vale la pena destacarlo para la aplicación de la indagación en ciencias en la escuela, es que los que participan en la indagación no conocen la respuesta a la pregunta o problema objeto de estudio, consideran importante investigarlo y están entusiasmados tratando de encontrar una respuesta o solución.

Plantear preguntas destaca el hecho que los estudiantes están comprometidos en contestar preguntas de verdadero interés para ellos, las cuales han estimulado su curiosidad. A menudo estas preguntas serán planteadas por el profesor o la profesora, otros estudiantes o emergerán de la lectura, pero, cualquiera sea el origen de la pregunta, en la indagación los alumnos deben tomarlas como propias, involucrando su curiosidad y el deseo de entender. Plantear y responder preguntas se equipara a veces con la resolución de problemas, donde la atención se centra en la búsqueda de una solución que "funcione". Sin embargo, en la ciencia la solución única no es suficiente. El desarrollo de teorías y modelos para explicar fenómenos requiere que las ideas sean "evaluadas en relación con las explicaciones alternativas y comparadas con la evidencia ... De esta manera entendiendo por qué la respuesta equivocada está equivocada puede ayudar a asegurar una comprensión más profunda y más sólida sobre por qué la respuesta correcta es correcta" ¹³.

11 National Science Foundation (NSF) (1997) *The Challenge and Promise of K-8 Science Education Reform*. Foundations, 1. Arlington, VA: NSF p7

12 IAP (2012) Taking Inquiry-Based Science Education into Secondary Education. Report of a global conference. <http://www.sazu.si/files/file-147.pdf>

13 National Research Council (2012) *A Framework for K-12 Science Education*. Washington DC: National Academies Press. p44

La discusión de las definiciones deja claro que el aprendizaje de la ciencia a través de la indagación es un proceso complejo en el cual se vinculan interactivamente el conocimiento y la comprensión y las habilidades de recoger y utilizar la evidencia. Las habilidades que son esenciales para la construcción de comprensión son tanto habilidades físicas como mentales, relacionadas a la generación de evidencia y al uso de evidencia para probar ideas que puedan ayudar a explicar un hecho o fenómeno que está siendo estudiado.

Al mismo tiempo, el uso de las habilidades involucra conocimiento y comprensión, no sólo saber cómo generar, recopilar e interpretar datos, sino también la comprensión de por qué es importante trabajar científicamente. Además, hay un elemento afectivo en el proceso, que influye en la voluntad de participar en las distintas acciones involucradas en llevar a cabo una indagación y advertir los resultados que pueden requerir un cambio en las ideas preexistentes. Todo esto, también se inserta en un contexto cultural que puede promover o inhibir el desarrollo de la comprensión a través de la indagación.

El reconocimiento de esta interdependencia entre los conocimientos y las habilidades ha llevado a sugerir que la indagación está mejor especificada en términos de *prácticas*, conjuntos complejos de acciones que llevan a experimentar y entender la ciencia como "un conjunto de conocimientos arraigados en evidencia".¹³ Sin embargo, en este libro seguiremos utilizando la palabra "habilidad" y "competencia" de manera intercambiable, siguiendo la convención de la OECD:

En el contexto de la Estrategia para las Habilidades de la OECD, los conceptos de "habilidad" y "competencia" se utilizan indistintamente. Por habilidad (o competencia) se entiende: el conjunto de conocimientos, atributos y capacidades que le permiten a una persona llevar a cabo con éxito y en forma consistente una actividad o tarea, ya sea concebida en forma amplia o acotada, y que pueda ser desarrollada y ampliada a través del aprendizaje.¹⁴

Fundamentos

El aprendizaje basado en la indagación es complejo y no es una opción fácil. Nos esforzamos para ponerlo en práctica porque creemos que promueve la comprensión y el desarrollo de las habilidades que necesitan los estudiantes para cumplir con las exigencias de la vida del siglo XXI. Es un hecho ampliamente aceptado¹⁵ que la educación científica debería permitir a los estudiantes desarrollar los conceptos claves de ciencias (grandes ideas) que les permitan comprender los acontecimientos y fenómenos de relevancia en sus vidas actuales y futuras. Los estudiantes también deben lograr comprensión sobre cómo se obtienen las ideas y el conocimiento científico y las habilidades y actitudes involucradas en la búsqueda y la utilización de la evidencia.

En el futuro, los jóvenes tendrán que elegir entre más opciones que aquellos que vivieron en las décadas pasadas. Deberán desarrollar las habilidades, la voluntad, la flexibilidad de pensamiento y la energía necesarias para tomar decisiones efectivas. La capacidad de seguir aprendiendo durante toda la vida se reconoce como esencial para las generaciones futuras y, por lo tanto, tiene que ser una característica de la educación de los estudiantes de todos los países, como subraya la OECD:

Los estudiantes no pueden aprender en la escuela todo lo que necesitarán saber en la vida adulta. Lo que deben adquirir son los requisitos previos para el aprendizaje exitoso en la vida futura. Estos requisitos son tanto de naturaleza cognitiva, como motivacional. Los estudiantes deben ser capaces de organizar y regular su propio aprendizaje, para

14 OECD (2011) *Towards an OECD Skills Strategy*. Paris: OECD. P7 footnote

15 OECD (2003), *The PISA 2003 Assessment Framework* Paris: OECD p132

Harlen (Ed) (2010) *Principles and Big Ideas of Science Education*.

aprender tanto de manera independiente como con otros, y para superar las dificultades en el proceso de aprendizaje. Esto requiere que ellos estén conscientes de sus propios procesos de pensamiento, las estrategias de aprendizaje y los métodos.¹⁶

Además, existe un amplio reconocimiento de la importancia de desarrollar las habilidades, actitudes, conocimientos y comprensión, las que se consideran más importantes que la acumulación de grandes cantidades de conocimiento factual. El conocimiento del contenido se puede encontrar fácilmente a partir de fuentes de información ampliamente disponibles a través del uso de computadores y especialmente el internet. Lo que los alumnos necesitan son los conocimientos necesarios para acceder a estas fuentes y la comprensión para seleccionar lo que es relevante y darle sentido.

Aprender es una actividad social en la cual el lenguaje juega un papel clave. La interacción con otros a menudo significa que los individuos llegan a un entendimiento compartido de ideas que podrían no haber alcanzado por sí solos. Las ideas que los estudiantes se forman a partir de la experiencia directa deben ser comunicadas y esto implica el uso de palabras que transmiten un significado a los demás. El proceso de expresar ideas a través del habla o la escritura, a menudo significa que las ideas tienen que ser reformuladas de manera que se ven influidas por el significado que otros les dan a las palabras. También es necesario aprender que la ciencia utiliza las palabras con significados precisos diferentes de su uso común en el lenguaje cotidiano, y utiliza las matemáticas y otros símbolos abstractos a la hora de cuantificar las observaciones del mundo. Volveremos al papel del lenguaje y sobre todo de la conversación, en el Capítulo 5.

Objetivos

Si bien es importante que los alumnos aprendan a aprender y desarrollen las habilidades de la indagación, es necesario que haya un equilibrio entre el aprendizaje conceptual y el aprendizaje sobre cómo hacer para aprender. Aprender acerca de cómo responder a una pregunta no es suficiente por sí solo, la pregunta también tiene que ser respondida. Por otro lado, la búsqueda de la respuesta a una pregunta en particular no es suficiente, pues sólo preocupándonos de cómo fue respondida permitirá apoyar el aprendizaje en contextos nuevos.

En resumen, a través de su educación en ciencias los estudiantes deberían desarrollar:

- comprensión de las ideas científicas fundamentales
- comprensión de la naturaleza de la ciencia, la indagación científica, el razonamiento
- competencias científicas de obtención y uso de evidencia
- actitudes científicas, tanto actitudes dentro de la ciencia y hacia la ciencia
- habilidades que apoyan el aprendizaje durante toda la vida
- capacidad de comunicarse utilizando lenguaje y representaciones apropiadas, incluyendo lenguaje escrito, oral y matemático
- apreciación de la contribución de la ciencia a la sociedad y de cómo la ciencia se utiliza en la tecnología y la ingeniería.

Un enfoque indagatorio, si se lleva a cabo eficazmente, ofrece la promesa de lograr estos objetivos en mayor grado que los enfoques tradicionales de la enseñanza y el aprendizaje de las ciencias.¹⁷ La reserva crítica aquí es si se "lleva a cabo eficazmente". La complejidad de ECBI, como se señaló anteriormente, hace que esto sea un desafío considerable. La aplicación puede necesitar un cambio

16 OECD (2000) *Measuring Student Knowledge and Skills: A new Framework for Assessment*. Paris: OECD. p90

17 Minner, D.D., Levy, A. J, and Century, J. (2010) Inquiry-Based Science Instruction—What Is It and Does It Matter? Results from a Research Synthesis Years 1984 to 2002, *Journal of Research in Science Teaching*, 47 (4) 474-496

fundamental en varios aspectos de la pedagogía, desde la disposición del espacio de aprendizaje (de modo que los estudiantes puedan trabajar en colaboración) a las preguntas que los profesores realicen, la retroalimentación que dan a los estudiantes y la naturaleza de su interacción con ellos, así como la interacción de los estudiantes con los objetos y fenómenos que investigan. El grado de cambio que puede ser requerido, se puede visualizar comparando las acciones de los estudiantes que participan en el aprendizaje basado en la indagación que se listan en el Recuadro 3, con aquellas de la enseñanza de la ciencia por transmisión en el Recuadro 4.

La enseñanza como "transmisión de hechos" fue el modo predominante cuando el principal objetivo de la educación científica era proporcionar a los futuros científicos los conocimientos esenciales, en lugar de tener también el propósito de ofrecer a todos la oportunidad de alcanzar los objetivos mencionados anteriormente. La ciencia se transmitía a los alumnos "lista para su uso" (*ready-made*), como opuesta a algo que se experimenta, donde el conocimiento es creado a través de la acción. Pero esto no quiere decir que la indagación sea la única forma de pedagogía que los estudiantes encuentran en su educación científica. Hay algunas cosas como las destrezas para el uso de equipamiento, nombres, convenciones y símbolos que pueden enseñarse mejor directamente y habrá ocasiones en las que la indagación contribuya a dar sentido a la experiencia, sin ser el único método utilizado. Sin embargo, cuando el objetivo es la comprensión, la indagación tiene un rol clave en la educación científica de los estudiantes.

Recuadro 3: Actividades de los estudiantes: aprendiendo a través de la indagación

- Los estudiantes abordan preguntas que han identificado como propias aunque hayan sido introducidas por el (la) profesor(a)
- No conocen la respuesta a las preguntas que investigan
- Saben lo suficiente sobre el tópico como para involucrarse en la pregunta
- Realizan predicciones basadas en sus ideas emergentes sobre el tópico
- Participan en la planificación de la investigación para someter a prueba sus predicciones
- Ellos mismos conducen las investigaciones
- Utilizan fuentes y métodos apropiados para recoger los datos relevantes para someter a prueba sus predicciones
- Discuten los que encuentran en relación con sus expectativas iniciales o sus predicciones
- Sacan conclusiones y tratan de explicar lo que encuentran
- Comparan sus hallazgos y sus conclusiones con lo que otros han encontrado y concluido
- Toman notas y hacen registros durante su trabajo
- Se involucran en discusiones de los métodos usados y de los resultados de las investigaciones

Recuadro 4: Actividades de los estudiantes: aprendiendo a través de transmisión

- Las actividades de los estudiantes siguen una secuencia establecida en un libro o por el(la) profesor(a) sin mayor preocupación de colocar lo que realizan en contexto
- Es posible que lean sobre como conducir una investigación, pero tienen escasa oportunidad de experimentar por ellos mismos el proceso
- Pueden observar demostraciones por parte del profesor pero es posible que no entiendan las razones de lo que se realiza
- Cuando realizan actividades prácticas siguen instrucciones con poca participación en decidir qué es lo que se hace
- Los experimentos que observan o conducen esta diseñados para confirmar una conclusión que ya se conoce; "experimentos para mostrar que ..."
- No siempre entienden por qué deben realizarse ciertos pasos en un experimento o investigación
- Dan cuenta de los resultados de las investigaciones en forma estructurada, a menudo copiadas de un libro o dictadas por el profesor o profesora
- Registran la "respuesta correcta" aún si no observaron lo que debería haber ocurrido
- Trabajan en forma independiente o en parejas y son desalentados de discutir su trabajo

Capítulo 3

Propósitos y usos de la evaluación del aprendizaje

Como se señaló en el Capítulo 1, toda evaluación del aprendizaje involucra la generación, recolección, interpretación y comunicación de datos. El proceso es similar si el propósito es ayudar al aprendizaje o resumirlo e informarlo. En este capítulo se exploran los dos objetivos principales de la evaluación - para ayudar al aprendizaje o para resumirlo e informar - y consideramos la viabilidad y la conveniencia de utilizar los datos recogidos con un cierto propósito para un propósito distinto. El que una determinada práctica evaluativa sea evaluación formativa o evaluación sumativa está esencialmente determinado por el *uso* que se hace de los datos. Así, por ejemplo, un test de aula puede ser utilizado para ayudar a los estudiantes y profesores a identificar lo que los estudiantes ya saben sobre un tema nuevo - un uso formativo - o para juzgar e informar sobre lo que han aprendido al final - un uso sumativo. Así más correctamente deberíamos referirnos al "uso formativo de la evaluación" y al "uso sumativo de la evaluación", sin embargo, la convención y conveniencia permiten el uso de los términos más breves que usamos aquí. Después de examinar la naturaleza y la importancia de la evaluación para estos dos propósitos, se considera la relación que existe entre ellos. La discusión de esta relación se continúa, en el Capítulo 6, después de revisar algunos procedimientos para la implementación de la evaluación formativa y sumativa.

Propósitos y usos

Una pregunta clave de toda evaluación es: ¿Cuál es su objetivo principal?¹⁸ En general se acepta que hay dos respuestas a esta pregunta:

- para ayudar a los estudiantes mientras están aprendiendo
- para averiguar lo que han aprendido en un momento determinado

Estas corresponden a los propósitos formativos y sumativos de la evaluación.

La evaluación formativa tiene el propósito de apoyar el aprendizaje y por eso también se le llama "evaluación *para* el aprendizaje" (AfL, sigla en inglés para *assessment for learning*). Involucra procesos de "búsqueda e interpretación de evidencia por parte de los aprendices y profesores para decidir dónde están los aprendices, hacia donde necesitan ir y de qué mejor forma llegar allí"¹⁹.

La evaluación sumativa tiene el propósito de resumir y reportar lo que se ha aprendido en un determinado momento y por esta razón también se le llama "evaluación *de* aprendizaje" (AoL, sigla en inglés para *assessment of learning*). Se trata de resumir a través de la revisión del aprendizaje durante un período de tiempo, y/o de monitorear el aprendizaje a través de pruebas en un momento determinado.

En el caso de la evaluación formativa existe un uso principal de los datos - ayudar al aprendizaje. Si la información sobre el aprendizaje de los estudiantes no se utiliza en las decisiones destinadas a ayudar ese aprendizaje, entonces el proceso no puede ser descrito como evaluación formativa. Por el contrario, los datos de la evaluación sumativa se pueden utilizar de varias maneras, algunas

18 Stobart, G. (2008) *Testing Times. The uses and abuses of assessment*. London: Routledge

19 Assessment Reform Group (ARG) (2002) *Assessment for Learning: 10 Principles*. www.assessment-reform-group.org

relacionadas con los estudiantes individuales y algunas para producir resultados agregados para grupos o poblaciones, no todos los cuales son usos adecuados o válidos. Como se señaló en el Capítulo 1, la validez no es una propiedad de un instrumento de evaluación o procedimiento en particular, sino que depende de cómo se utiliza y de las inferencias extraídas del resultado de su uso. Los resultados de una prueba de recuerdo de conocimiento, no indican logro a través de todo el dominio de la materia - una prueba de aritmética no se debe utilizar para indicar el desempeño en matemáticas, ni debe ser utilizada como una medida de la calidad de la enseñanza. Aunque estos puntos puedan parecer obvios, es un hecho que este abuso del uso de los datos de pruebas se produce. Por ejemplo, Newton²⁰ identificó 16 usos de los resultados de las pruebas nacionales en Inglaterra. Estos iban desde la evaluación de programas, la fijación de objetivos para los estudiantes y las escuelas, el monitoreo de las escuelas y de los estudiantes, la selección de los estudiantes, hasta la elección de colegios por los padres, e incluso la valoración de la propiedad en las áreas donde hay escuelas con puntajes altos o bajos. En algunos casos es legítimo usar medidas del logro de los estudiantes como parte de los datos para hacer juicios (por ejemplo, como un elemento en la evaluación de la escuela), pero usarlo como única medida - y en particular cuando se les asocian recompensas y sanciones - invita acciones inapropiadas para inflar el desempeño medido. Volveremos a esto en el Capítulo 4, donde se considera el uso de datos agregados para la evaluación de logros de la escuela y del sistema, y el establecimiento de metas.

Ahora nos preocuparemos, más en detalle, de estos dos objetivos principales de la evaluación - formativo y sumativo - discutiendo en cada caso qué es y por qué es importante. Los Capítulos 5 y 6 se preocupan de los métodos de implementación.

Evaluación formativa

Qué es

Algunas de las diversas definiciones de la evaluación formativa que se han propuesto en las últimas dos décadas han sido revisadas por Wiliam²¹ quien sugiere que las características principales se pueden reunir en la siguiente definición:

La práctica en el aula es formativa en la medida en que la evidencia sobre los logros de los estudiantes es provocada, interpretada y utilizada por los profesores, los aprendices, o sus compañeros, para tomar decisiones sobre los próximos pasos en la instrucción, los que se espera sean mejores, o estén mejor fundados, que las decisiones que habrían tomado en ausencia de la evidencia que se obtuvo.²²

La Figura 1²³ representa este proceso como un ciclo de eventos. La evaluación formativa no es algo que sucede de vez en cuando; sino que es parte integral del proceso de toma de decisiones que está sucediendo todo el tiempo en la enseñanza. Las actividades representadas por A, B y C se dirigen hacia los objetivos de la lección, o serie de lecciones sobre un tema. Estos objetivos, compartidos con los alumnos por parte del profesor o la profesora, se expresan en términos *específicos*, por ejemplo en una clase de ciencias puedan ser "planificar y llevar a cabo una investigación sobre las condiciones preferidas por los chancitos de tierra (o cochinillas de humedad)". El trabajo de los alumnos en la actividad A, dirigida a los objetivos, brinda la oportunidad, tanto a el(la) profesor(a) como a los estudiantes, de obtener evidencia del progreso hacia las metas.

20 Newton, P.E. op cit pp270-272

21 Wiliam, D. (2009) An integrative summary of the research literature and implications for a new theory of formative assessment, in (eds) H. L. Andrade and G. J. Cizek, *Handbook of Formative Assessment*, New York: Taylor and Francis

22 Black, P. and Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21 (1), 5-13. p9

23 Adapted from Harlen, W. (2006) *Teaching, Learning and Assessing Science 5-12*. London: Sage. p 87.

Con el fin de interpretar la evidencia, en este ejemplo, tanto el(la) profesor(a) como los estudiantes necesitan saber qué significa "buena planificación", por lo que los estudiantes necesitan tener una cierta comprensión de los criterios a aplicar en la evaluación de su trabajo (¿Está tomando en cuenta todas las variables la investigación planificada? ¿Qué evidencia se recogerá y cómo?). El juicio lleva a la decisión sobre los próximos pasos relevantes que pueden ser intervenir o simplemente seguir adelante. Como Wiliam señala, "la evaluación formativa no requiere modificar la instrucción para ser formativa - simplemente puede confirmar que el curso de acción propuesto es de hecho el más apropiado".²⁴ La actividad B es el resultado de esta decisión y la fuente de evidencia para un siguiente ciclo de producción e interpretación de evidencia.

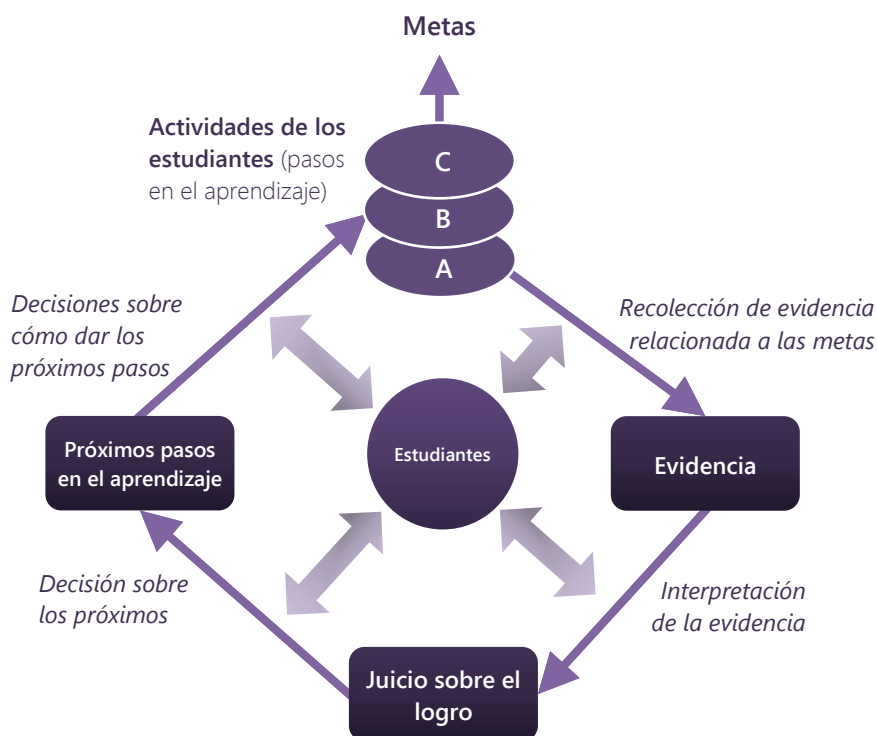


Figura 1:
Evaluación con propósitos formativos (adaptado de Harlen, 2006)

Los estudiantes están en el centro del proceso, ya que son ellos los que "realizan" el aprendizaje. Las flechas de dos cabezas que vinculan a los estudiantes a las distintas partes del ciclo de evaluación indican que los estudiantes reciben comentarios del profesor o la profesora, así como también proporcionan información. Ellos participan en decisiones, cuando esto resulta apropiado, a través de la auto- y co-evaluación.

En la evaluación formativa, los juicios sobre el progreso y las decisiones sobre los próximos pasos toman en cuenta las circunstancias, el aprendizaje pasado y el esfuerzo de alumnos individuales, así como lo que son capaces de hacer en relación con los objetivos del trabajo en un momento determinado. Así, los juicios deben estar tanto referidos a los estudiantes, como a los criterios (véase el Capítulo 1). Este enfoque apoya el aprendizaje mucho más que la simple aplicación de las mismas

²⁴ Leahy, S. and Wiliam D. (2012) From teachers to schools: scaling up professional development for formative assessment, in (ed) J. Gardner *Assessment and Learning*. London: Sage. 49-71, p 51.

normas a todos los estudiantes, lo que sería muy desmotivador para los estudiantes de bajos logros. En el caso de la evaluación formativa no se hacen comparaciones entre los estudiantes.

Las acciones indicadas por las flechas en la Figura 1 no son "etapas" en una lección, ni necesariamente el resultado de decisiones planificadas realizadas por el(la) profesor(a). Ellas representan el pensamiento involucrado en focalizarse en qué y en cómo están aprendiendo los estudiantes y el uso de esto para ayudar al aprendizaje que sigue. En algunos casos, puede ser posible que los profesores y los estudiantes decidan juntos sobre una acción inmediata. En otros casos, el profesor o la profesora pueden tomar nota de qué ayuda es necesaria y proveerla en un momento posterior.

Implementar la evaluación formativa significa que no todo en una lección se puede planificar por adelantado. Por definición, si las ideas previas de los estudiantes deben ser tomadas en cuenta, algunas de las decisiones dependerán de lo que sean estas ideas. Algunas ideas pueden ser anticipadas a partir de la experiencia de los profesores y de los resultados de investigación integrados en los recursos curriculares, pero no todas. Lo que el(la) profesor(a) necesita no es el contenido prescrito para la lección, sino un conjunto de estrategias a desplegar de acuerdo con lo que se considera apropiado, en ocasiones particulares. Algunos ejemplos de estas estrategias se presentan en el Capítulo 5.

Recuadro 5: Prácticas claves de la evaluación formativa

- Los estudiantes están involucrados en expresar y comunicar su comprensión y habilidades a través de diálogo en el aula, iniciado por preguntas abiertas y centradas en la persona
- Los estudiantes comprenden los objetivos de su trabajo y qué es un trabajo de buena calidad
- Los estudiantes reciben retroalimentación sobre cómo mejorar o avanzar y se evita hacer comparaciones con otros estudiantes
- Los estudiantes participan en la autoevaluación de manera que toman parte en la identificación de lo que necesitan hacer para mejorar o avanzar
- Existe diálogo entre el(la) profesor(a) y los estudiantes que fomenta la reflexión sobre su aprendizaje
- Los profesores utilizan la información sobre el aprendizaje que está en curso para ajustar la enseñanza de manera que todos los estudiantes tengan la oportunidad de aprender²⁵

La retroalimentación es una característica esencial de la evaluación formativa. La retroalimentación de doble vía, del profesor a los alumnos y de los alumnos al profesor, implica una visión del aprendizaje como un proceso en el que la comprensión se construye activamente por parte de los estudiantes

- La retroalimentación del profesor a los alumnos les ofrece información que les ayuda a tomar los próximos pasos para mejorar su comprensión o habilidades. La forma y el foco de la retroalimentación tiene que ser cuidadosamente sopesada por el profesor o la profesora. El foco de la retroalimentación influye en aquello a lo cual los estudiantes prestan atención y la forma que adopta determina si se puede utilizar para avanzar en el aprendizaje. En una perspectiva del aprendizaje en la cual el aprendizaje se equipara con "ser enseñado", la retroalimentación al estudiante desde el(la) profesor(a) es sobre la calidad o el éxito del trabajo de los estudiantes

más que en cómo mejorarlo. La evaluación formativa realmente no tiene ningún papel en el aprendizaje visto de esta manera.

- La retroalimentación hacia la enseñanza, que ocurre desde los estudiantes a los profesores, es necesaria para que los docentes puedan ajustar los desafíos que ofrecen a los estudiantes de manera que no sean demasiado exigentes, colocando el éxito fuera de alcance, ni demasiado simples para que resulten atractivos. Utilizando la retroalimentación de las observaciones de los estudiantes y su trabajo para juzgar la capacidad de los estudiantes para tomar ciertos pasos con ayuda (la zona de desarrollo potencial, como se discute en el Capítulo 4) es una tarea compleja y difícil para los profesores. Muchos profesores necesitan una buena dosis de ayuda con esta tarea si van a usar la retroalimentación para regular la enseñanza con el fin de optimizar el aprendizaje.

En resumen, los componentes claves de la evaluación formativa se resumen en el Recuadro 5²⁵. En el Capítulo 5 se considera cómo estos componentes de la evaluación formativa pueden ser puestos en práctica

Por qué es importante

La importancia de la evaluación formativa radica en la evidencia de su efectividad para mejorar el aprendizaje. Los estudios empíricos de evaluación en el aula han sido objeto de varias revisiones de investigación. La revisión de Black y Wiliam (1998) atrajo la atención mundial debido, en parte, al intento de cuantificar el impacto del uso de la evaluación formativa. Desde entonces ha habido una serie de otras revisiones e investigaciones que han justificado las afirmaciones realizadas por Leahy y Wiliam:

La conclusión general es que para un rango amplio de diferentes asignaturas, en distintos países, y para alumnos de distintas edades, el uso de la evaluación formativa parece estar asociada con mejoras considerables en la tasa de aprendizaje. Es difícil estimar cuán grandes son estas ganancias... pero parece razonable concluir que el uso de la evaluación formativa puede aumentar la tasa de aprendizaje de los estudiantes en un 50% a 100%.²⁶

Stobart,²⁷ sin embargo, pone una nota de cautela, señalando que, aparte de un estudio realizado por Wiliam *et al* (2004) sobre el impacto de su proyecto de investigación-acción en el logro de los estudiantes, "existe, hasta ahora, poca evidencia empírica directa del impacto [de la evaluación formativa] en el logro". Señala que la mayoría de los estudios de evaluación se han focalizado en la magnitud de cambio en la práctica de los profesores y en las actitudes de los estudiantes y su participación, más que en el aprendizaje conceptual de los alumnos. Sin embargo, se puede argumentar que tales cambios son pasos necesarios para la mejora del aprendizaje. Más aún, el número de influencias sobre el aprendizaje medido de los alumnos, más allá de cambios más bien sutiles en la pedagogía cuando se lleva a cabo la evaluación formativa, hacen que su impacto sea difícil de detectar. De hecho, Wiliam *et al* (2004) señalan que las comparaciones en que ellos basan sus afirmaciones no son todas «igualmente robustas».

La importancia de la evaluación formativa para ECBI radica en la intención que tienen ambas de apoyar el desarrollo de una comprensión verdadera y de las competencias necesarias para el aprendizaje continuo. La enseñanza para el desarrollo de la comprensión involucra tomar en cuenta las ideas y las habilidades previas de los estudiantes y promover la progresión mediante el ajuste del

25 Harlen, W. (2007) op cit p121

26 Leahy, S. and Wiliam, D. (2012) op cit p52

27 Stobart, G. (2008) op cit p154

desafío para que coincida con estas ideas iniciales.²⁸ La práctica de la evaluación formativa, a través de profesores y estudiantes recogiendo datos sobre el aprendizaje mientras éste se lleva a cabo y la retroalimentación de información para regular el proceso de enseñanza y aprendizaje, está claramente alineada con los objetivos y la práctica del aprendizaje basado en la indagación. También apoya la apropiación por parte del alumno de su aprendizaje a través de la promoción de la autoevaluación y la participación en las decisiones sobre los próximos pasos, lo que ayuda a los estudiantes a asumir cierta responsabilidad por su aprendizaje en la escuela y fuera de ella.

Evaluación sumativa

Ya que acabamos de describir la evaluación formativa como teniendo un papel positivo en el aprendizaje, existe una tendencia a considerarla como la "buena" cara de la evaluación, y a la evaluación sumativa, que tiene un rol diferente, como la "mala" cara. Esto es lamentable, en varios sentidos. En primer lugar, aun cuando la evaluación sumativa no está diseñada para tener un impacto directo en el aprendizaje mientras éste se lleva a cabo, como es el caso de la evaluación formativa, también puede utilizarse para apoyar el aprendizaje de una manera menos directa, pero necesaria; por ejemplo, cuando se proporciona un resumen de los aprendizajes de los estudiantes para informar al próximo profesor(a) cuando los estudiantes pasan de una clase o escuela a otra. En segundo lugar, permite a los profesores, los padres y las escuelas realizar un seguimiento del aprendizaje de los estudiantes, tanto como individuos y como miembros de ciertos grupos (e.g. los de alto logro o los que necesitan ayuda especial). Tercero, proporciona datos que, junto con factores de contexto, pueden ser utilizados para la evaluación y la mejora de las escuelas. La mala reputación de la evaluación sumativa se debe a un uso indebido de datos que no reflejan plenamente los objetivos de aprendizaje. El peligro es severo en el caso de ECBI por su gran gama de diferentes objetivos, que no son fácilmente evaluados por los métodos convencionales. Esto hace que sea aún más importante tener en cuenta las formas en que se puede reunir información confiable para la evaluación sumativa.

Qué es

La evaluación sumativa es el nombre dado a la evaluación que se lleva a cabo con el propósito de informar los logros en un momento particular. Puede tener, y a menudo es el caso, algún impacto en el aprendizaje, y el resultado puede ser utilizado en la enseñanza, pero esa no es su razón principal. Volveremos a la relación con la evaluación formativa más adelante, pero por el momento se representará el proceso como proporcionando información exclusivamente para informar los logros como en la Figura 2.

La evidencia deriva de los tests, tareas especiales o actividades regulares y puede ser recogida por una serie de medios desde diferentes fuentes: respuestas escritas, artefactos construidos por los estudiantes, portafolios, observación de acciones, discusión o presentaciones de trabajo. Es claro que la obtención de evidencia sobre el logro en relación a toda la comprensión y las competencias

28 Bransford, J.D., Brown, A. and Cocking, R.R. (Eds) (2000) *How People Learn, Brain, Mind, Experience and School*. Washington, D.C.: National Academy Press.

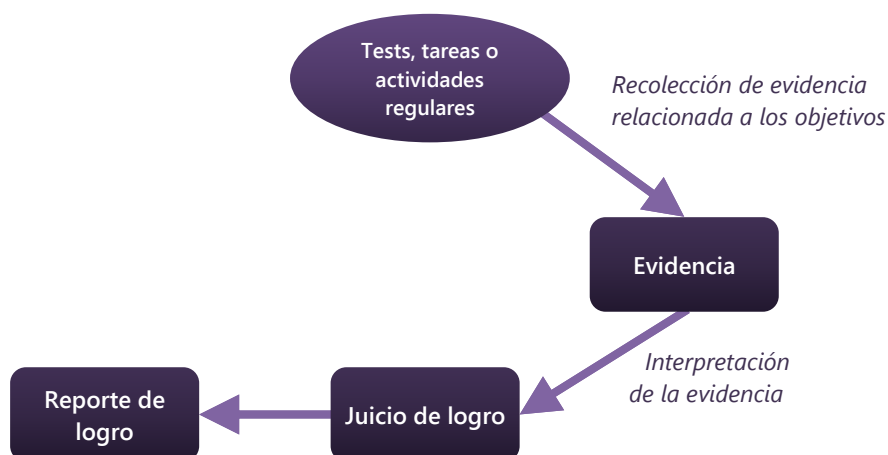


Figura 2:
Evaluación para propósitos sumativos (adaptado de Harlen, 2006)

pertinentes es la parte más importante del proceso, pues sin ella es poco probable que el informe final sobre el desempeño proporcione información confiable sobre el logro de los alumnos respecto a los objetivos de aprendizaje. Los pros y los contras de la utilización de diferentes fuentes de datos se analizan en el Capítulo 6.

La evidencia es interpretada en comparación con criterios o estándares relacionados a objetivos generales, más que a los objetivos de lecciones o temas específicos, como en el caso de la evaluación formativa. La nota o puntaje puede ponerla el (la) profesor(a) o un agente externo, como ocurre en algunas pruebas y exámenes nacionales. Sólo en las pruebas de aula más informales los estudiantes tienen comúnmente un papel en el proceso. Todos los estudiantes son juzgados por los mismos criterios, o esquemas de asignación de nota (rúbricas), mientras que, como se señaló anteriormente, en la evaluación formativa el criterio puede ser ipsativo, o referido al estudiante (véase el Capítulo 1) con el fin de ayudar a los estudiantes a reconocer sus avances desde distintos puntos de partida.

La interpretación necesariamente reduce la riqueza del desempeño propiamente tal a un puntaje, o nota que lo representa, por lo que se pierde una gran cantidad de información. Dependiendo del uso que se haga del resultado, el proceso de interpretación podrá incluir algún procedimiento para aumentar la fiabilidad. Cuando los resultados se utilizan para comparar a los estudiantes, sobre todo cuando está en juego la selección o calificación de alto impacto, se toman medidas para controlar la asignación de puntaje y moderar los juicios de los profesores o examinadores. Cuando la evaluación sumativa es esencialmente basada en el aula y está en las manos de los profesores existe la posibilidad de que la evidencia sea recogida y utilizada sobre una amplia gama de tipos de logros.

La ausencia de referencia a los estudiantes en la Figura 2 reconoce que en general no tienen un papel en la evaluación sumativa. Sin embargo, cuando el proceso es abierto y los criterios de evaluación son compartidos con los estudiantes y los usuarios de los resultados, y no se limitan a lo que se puede hacer en una situación controlada o de prueba, hay una oportunidad para que los estudiantes tengan un papel en el proceso, por ejemplo, en la selección de los ítems en un portafolio. Queda, por supuesto, la obligación de garantizar que los juicios son confiables y se basan en los mismos criterios para todos los estudiantes. Consideramos maneras de hacer esto en el Capítulo 6.

Recuadro 6: Prácticas claves de la evaluación sumativa

- Los estudiantes se involucran en tareas especiales o tests como parte de, o además del trabajo regular
- Ocurre en ciertos momentos cuando el logro requiere ser informado, no como un ciclo que es parte regular del aprendizaje
- Se relaciona al logro de objetivos amplios expresados en términos generales, más que a los objetivos de actividades de aprendizaje particulares
- Involucra el logro de todos los estudiantes siendo juzgados contra los mismos criterios o esquema de calificación
- Requiere de algunas medidas para asegurar la confiabilidad
- Ofrece oportunidades limitadas para la autoevaluación de los estudiantes

La forma del reporte depende en gran medida de la naturaleza de la tarea, las bases para el juicio y la audiencia a la que se reporta. Las calificaciones numéricas de las pruebas son una sumatoria de un conjunto diverso de preguntas. El mismo total se puede lograr de muchas maneras, por lo que los puntajes tienen poco significado respecto a lo que los estudiantes saben o pueden hacer. También dan una falsa impresión de precisión, lo que está muy lejos de ser el caso. Los puntajes pueden usarse directamente para ranquear a los estudiantes, pero esto sólo es útil en el contexto de la selección ya que una posición de ranking no da ninguna indicación con sentido en términos de aprendizaje.

En teoría, la presentación de informes contra criterios que describen el desempeño en niveles progresivos o grados, puede proporcionar una indicación más significativa de lo que los estudiantes han logrado (ver Capítulo 6). Sin embargo, a fin de preservar un cierto significado en el informe, un perfil es preferible a un solo grado global o nivel que tendría que combinar diferentes dominios. La abreviatura de "niveles" - las etiquetas dadas a criterios progresivos - puede ser útil para algunos propósitos, pero para informar a los padres y estudiantes los niveles deben ser explicados o acompañados por descripciones de lo que el estudiante puede hacer. Además, como se señala más adelante (Capítulo 6, página 80), el uso de niveles puede tener consecuencias negativas para la motivación de los alumnos y el aprendizaje.

Algunas prácticas características de la evaluación sumativa se resumen en el Recuadro 6.

Por qué es importante

Se han dado varias razones al comienzo de esta sección. La más convincente, sin embargo, es que la evaluación sumativa es importante porque es necesaria. Es inevitable - los informes sobre el aprendizaje de los estudiantes tienen que hacerse y deben llevarse registros a intervalos regulares. Por el contrario, la evaluación formativa puede considerarse, en cierto sentido voluntaria, ya que es posible enseñar sin ella. Sin embargo, los profesores tienen que mantener registros que resuman el desempeño de los estudiantes en ciertos momentos claves, como al final de tópicos o semestres, y utilizar estos registros en su planificación. Los padres y los próximos profesores de los estudiantes en los puntos de transición de una clase a otra, o de una escuela a otra, necesitan registros de lo que han logrado. Los directores y gerentes de las escuelas deben contar con registros para examinar el progreso de grupos de estudiantes durante su paso por la escuela para usarlo en la autoevaluación de la escuela y la planificación curricular.

También es importante porque *de facto* lo que se evalúa es tomado como una señal de aprendizaje importante. Por desgracia, es a menudo el caso que aquello que se evalúa es lo que *se puede* evaluar en lugar de lo que *debe ser* evaluado. La diferencia entre éstos es probable que sea

particularmente importante en el caso de ECBI, donde los objetivos se refieren a la construcción de comprensión y al desarrollo de "habilidades utilizadas por los científicos. En el Capítulo 6 se consideran los "pros y contras" de los distintos métodos de evaluación sumativa, pero vale la pena señalar aquí que no existe un enfoque o método "perfecto". Toda evaluación es sólo una muestra de lo que se ha aprendido y una aproximación de lo bien que se ha aprendido, y el resultado de varios juicios subjetivos. Una mejor comprensión del proceso de evaluación por todos los involucrados - desde los profesores y los estudiantes a los políticos y empresarios - podría ayudar a aflojar el control de la evaluación sobre el currículum. Volveremos a esto en el Capítulo 4.

La relación entre la evaluación formativa y sumativa

Una de las razones por la que la evaluación sumativa ha ganado la reputación de ser la cara mala de la evaluación es que cuando el desempeño medido se convierte en el factor dominante en el aula, desplaza la práctica de la evaluación formativa. La naturaleza de la retroalimentación dada por los profesores a los estudiantes no solo orienta los próximos pasos, sino que contribuye a la impresión general que los estudiantes tienen de la capacidad de sus profesores para ayudarlos y del interés en ellos como aprendices. Esto se aprecia en el siguiente informe de investigación:

Roderick y Engel²⁹ informaron sobre cómo una escuela que brinda un alto nivel de apoyo fue capaz de incrementar el esfuerzo y el desempeño en las pruebas de los estudiantes de muy bajos logros y escaso interés, en un mayor grado que una escuela comparable que entregaba un bajo nivel de apoyo a estudiantes similares. Mayor apoyo significaba crear un ambiente de apoyo social y educativo, trabajar duro para incrementar el sentido de auto-eficacia de los estudiantes, focalizándose en los objetivos relacionados al aprendizaje, haciendo explícitos los objetivos, utilizando la evaluación para ayudarles a tener éxito y creando mapas cognitivos que hacían evidentes los progresos. También demostraron un fuerte sentido de responsabilidad hacia sus estudiantes. Bajo apoyo del profesor o profesora significaba profesores que no veían los objetivos de aprendizaje como alcanzables, no estimulaban el trabajo a través de actividades significativas, no reconocían el cambio y la motivación por parte de los estudiantes, y no hacían conexiones personales con los estudiantes en relación con los objetivos como aprendizaje³⁰.

Pollard et al³¹ han señalado que la introducción de las pruebas nacionales en Inglaterra y la exigencia a los profesores de asignar niveles a los estudiantes, ha afectado su respuesta a los estudiantes y el uso de la evaluación formativa. Los estudiantes estaban conscientes de que si bien el esfuerzo se estimulaba, lo que contaba era el resultado en las pruebas. No es de extrañar, entonces, que la evaluación sumativa haya adquirido una mala reputación, a pesar de que a menudo el problema se deriva del uso que se hace de los resultados más que de la naturaleza de la evaluación sumativa en sí. Como hemos señalado, la evaluación sumativa es necesaria e inevitable, lo que lleva a la pregunta: ¿existen formas en que se pueda realizar y usar la evaluación sumativa sin tener un impacto perjudicial sobre la evaluación formativa?

Un enfoque obvio es pensar al revés, es decir, focalizarse en la evaluación formativa y ver si es posible obtener datos de este proceso formativo para proporcionar un juicio sumativo.

29 Roderick, M. and Engel, M. (2001) The grasshopper and the ant: motivational responses of low achieving pupils to high stakes testing. *Educational Evaluation and Policy Analysis* 23: 197-228

30 Harlen, W. (2012a) The role of assessment in developing motivation for learning, in (ed) J. Gardner *Assessment and Learning*, pp171-184. p177

31 Pollard, A., Triggs, P., Broadfoot, P., Mcness, E. and Osborn, M. (2000) *What pupils say: changing policy and practice in primary education* (chapters 7 and 10). London: Continuum

Tradicionalmente, el sistema educativo danés ha trabajado de esta manera. Al no tener pruebas nacionales en la enseñanza obligatoria, los profesores de los primeros grados proporcionan retroalimentación a los estudiantes sobre su progreso en un proceso continuo, y en los cursos superiores el(a) profesor(a) resume los progresos cada cuatro o seis meses en la forma de una nota. Al final de la enseñanza obligatoria, la nota más reciente aparece en el certificado. Para algunas materias se realiza un examen oral con un asesor externo, normalmente un profesor o profesora de una escuela vecina, y la nota de estos exámenes también se registra en el certificado final. Entre los problemas obvios está, cómo asegurar la fiabilidad y la protección de los estudiantes contra los profesores hostiles.

Las dos listas de características de la evaluación formativa (Recuadro 5) y sumativa (Recuadro 6) y las representaciones en las Figuras 1 y 2 muestran algunas diferencias claves en el uso que se hace de la evidencia. Sin embargo, como se insinuó anteriormente, existe el potencial de que la evidencia recogida para propósitos sumativos se retroalimente hacia la enseñanza y el aprendizaje, para identificar los aspectos en los que se necesita atención para mejorar el desempeño. Black et al.³² ofrecen ejemplos de profesores que utilizan pruebas en el aula para que los estudiantes puedan identificar sus puntos débiles y concentrar un mayor esfuerzo. En la práctica, este enfoque puede ser utilizado por los profesores principalmente en el contexto de las pruebas de aula sobre las que tienen todo el control. Mientras que algunas pruebas externas y los exámenes se pueden utilizar de esta manera, discutiendo los resultados con los estudiantes, existe el peligro de que el proceso pueda pasar del desarrollo de la comprensión a "enseñar para la prueba".

Un ejemplo de la combinación de los propósitos formativos y sumativos en la evaluación de alto impacto para los estudiantes, es el enfoque que ha sido utilizado por muchos años en el Certificado Superior en Queensland, Australia. Es importante reconocer que éste está diseñado tanto para fomentar la participación de los estudiantes, como para procedimientos de aseguramiento de la calidad. Un portafolio de evidencias, construido a lo largo de los dos últimos años de estudio, proporciona retroalimentación a los estudiantes permitiéndoles mejorar su desempeño durante este período, mostrando al mismo tiempo lo que han logrado al final. Maxwell explica que:

Para que este enfoque funcione, es necesario expresar las expectativas de aprendizaje en términos de dimensiones comunes de aprendizaje (criterios). Así puede haber discusión sobre si el estudiante ha alcanzado el resultado esperado y qué hay que hacer para mejorar el desempeño en futuras evaluaciones en las que aparecen las mismas dimensiones.

A medida que el estudiante construye el portafolio de evidencias de su desempeño, la evaluación anterior puede ser sustituida por la más reciente que cubre las mismas dimensiones de aprendizaje. El objetivo es dar a conocer "a donde ha llegado" el estudiante en su viaje de aprendizaje, y no cuál fue su punto de partida, o dónde estaba en promedio a lo largo de todo el curso.³³

Los criterios de evaluación se publican para que los estudiantes y padres de familia, tanto como los profesores, puedan estar familiarizados con ellos. Estos criterios describen lo que los estudiantes pueden hacer en diferentes categorías y subcategorías en cinco niveles o estándares (ver ejemplo en el Recuadro 20, Capítulo 6). La evidencia del portafolio se compara con los criterios usando juicios ("on-balance") de mejor ajuste. Dos condiciones claves para este enfoque son: tiempo para que los

32 Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. (2003). *Assessment for Learning: Putting it into Practice*. Maidenhead, England: Open University Press.

33 Maxwell, G. (2004) 'Progressive assessment for learning and certification: some lessons from school-based assessment in Queensland' Paper presents at the third conference of the Association of Commonwealth Examination and Assessment Boards, Redefining the Roles of Educational Assessment, March, Nidi, Fiji. p2-3

profesores participen en la moderación para garantizar la fiabilidad de los resultados y respeto por el profesionalismo de los profesores. También vale la pena señalar que el éxito de los estudiantes en el Certificado Superior está disociado de los procedimientos de rendición de cuentas de la escuela y de los profesores. Las escuelas, sin embargo, son estimuladas a utilizar los datos del proceso de certificación para la auto-evaluación y el mejoramiento escolar.

Estos ejemplos corresponden a casos de evaluación sumativa en los que la información se utiliza formativamente, que es una forma de ver la relación entre la evaluación para estos dos propósitos. Un corolario es el uso de la evidencia recopilada para la evaluación formativa siendo utilizada con fines sumativos. Dado que la evaluación formativa está en las manos de los profesores, se desprende que la evidencia que se utilizará en la evaluación sumativa será recogida por el(la) profesor(a), dando la oportunidad de reunir el rango más amplio de evidencia que proveen las actividades de clase. Un ejemplo de uso dual de los datos recogidos por los profesores se da en el Capítulo 6 (página 79).

Sin embargo, hay diferencias importantes que deben tenerse en cuenta sobre la forma en que se emiten los juicios cuando la evidencia se utiliza formativamente o sumativamente. Los juicios formativos tienden a ser referenciados al estudiante así como a los objetivos de la lección, mientras que los juicios de evaluación sumativa solo requieren ser referidos a criterios, utilizando criterios más amplios relacionados con objetivos de largo plazo. Se debe distinguir entre evidencia y juicio, de manera que las *evidencias* usadas en la evaluación formativa sean analizadas contra criterios más amplios que definan niveles o grados. Las medidas adicionales que se requieren para garantizar un uso confiable de la evaluación por profesores con fines sumativos, se discuten en el Capítulo 6.

Un último punto respecto a la relación entre evaluación formativa y sumativa es preguntarse si hay algún valor en hacer una distinción entre ellas, o si la relación debería considerarse más como una dimensión más que como una dicotomía.³⁴ Hay diferentes formas de utilizar la evaluación formativa, así como hay diferentes formas de recoger evidencias para la evaluación sumativa. Alguna de la evaluación formativa requiere una respuesta inmediata a las acciones de los estudiantes, mientras que en otros casos requiere algo de reflexión y planificación. También hay una diferencia entre si la recolección de evidencia surge como parte de la actividad de aprendizaje o se planifica con el fin de averiguar lo que se ha aprendido. La evaluación formativa planificada puede tener alguna similitud con la evaluación sumativa por los profesores. Lo que la haría formativa sería el uso que se haga de la información. Esto vuelve difusa la distinción entre evaluación formativa y sumativa y la relación sería tal vez mejor considerada como una dimensión más que una dicotomía. Sin embargo, la importancia de preservar la distinción radica en el rol de la evaluación como ayuda para el aprendizaje porque si no consideramos esto, entonces todo podría llegar a ser evaluación sumativa.

34 Harlen, W. (2012b) On the relationship between assessment for formative and summative purposes, in (ed) J. Gardner *Assessment and Learning*. London: Sage 87-102. pp 97-100

Capítulo 4

Evaluación del aprendizaje, pedagogía y currículum

Una de las principales razones de esta publicación es abordar las relaciones que deben establecerse entre la evaluación, el contenido del currículum y la pedagogía (Figura 3). Estas relaciones son de particular importancia en el contexto del desarrollo de la comprensión e implementación de la educación en ciencias basada en la indagación (ECBI). Los métodos tradicionales y el contenido de la evaluación rara vez reflejan los objetivos claves de ECBI. Por lo tanto, la evaluación tal como se practica corrientemente, tiende a actuar como un freno para la aplicación de la educación basada en la indagación. El desafío es cambiar las prácticas de evaluación, para que sirvan de apoyo a la enseñanza y el aprendizaje de las ciencias a través de la indagación y no, por el contrario, los limite.

En la primera sección de este capítulo, se consideran los posibles impactos positivos y negativos de la evaluación sobre aquello que se enseña y cómo se enseña. Una fuerte evidencia del impacto negativo surge del uso de los resultados de evaluaciones de los estudiantes para la evaluación con alto impacto de profesores y escuelas. La objeción a este uso se basa en la injusticia de utilizar los resultados de las pruebas como único indicador de la calidad de la enseñanza. Esto no quiere decir que los profesores y las escuelas no deban ser considerados responsables de los logros de sus estudiantes, sino que se debe tener en cuenta otro tipo de información, como se discute en la segunda sección. La tercera parte del capítulo aborda el tema de la perspectiva del aprendizaje en que se basa la evaluación tradicional, la que entra en conflicto con la visión del aprendizaje, implícito en la educación en ciencias basada en la indagación. Finalmente se argumenta que la evaluación, bien diseñada e implementada, tiene un rol en contribuir a que la educación cumpla con algunos de sus principales desafíos en relación con los problemas globales creados por la actividad humana.

Impactos de la evaluación

Las relaciones entre la evaluación, la pedagogía y el contenido se expresan frecuentemente en la forma de un triángulo, pero a menudo con el "currículum" en un vértice. Si tomamos "el currículum" en el sentido de toda la experiencia de los alumnos en la escuela, entonces cada uno de ellos es un aspecto del currículum como se ilustra en la Figura 3. Sin embargo, el uso más actual del término "currículum" se refiere implícitamente a los contenidos separándolos de la pedagogía y la evaluación. Aquí nos referimos al "contenido curricular" para dejar claro que el contenido es sólo un aspecto del currículo experimentado por los estudiantes, el todo incluyendo el contenido, la pedagogía y la evaluación y los supuestos subyacentes sobre el aprendizaje.

En la Figura 3, las flechas reconocen lo que es bien sabido - que lo que enseñamos es influenciado por la forma en que enseñamos, y el qué y cómo evaluamos influencia el qué y cómo enseñamos. Estas interacciones son importantes, ya que no sirve de nada abogar por el uso de la enseñanza basada en la indagación si hay una evaluación dominante (ya sea por medio de tests o juicios de los profesores) o un plan de estudios sobrecargado de contenido.

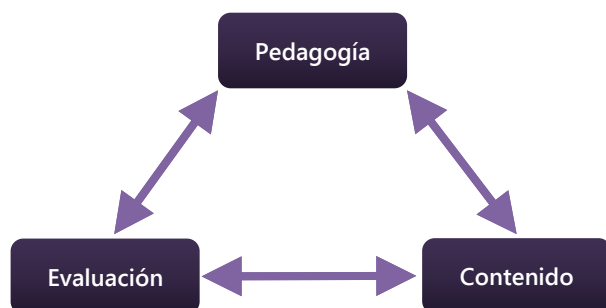


Figura 3:
Interacciones entre todos los aspectos del "currículum"

No sirve de nada sugerir que el contenido debe centrarse en las "grandes" ideas, si la evaluación requiere de la memorización de múltiples datos o si la pedagogía no establece vínculos que son necesarios para formar estas grandes ideas; es inútil esperar que los alumnos desarrollen responsabilidad por su propio aprendizaje continuo si la enseñanza no deja tiempo para la reflexión y el espacio para la creatividad. Tampoco podemos esperar actitudes positivas hacia la ciencia si el contenido del currículo resulta, a los ojos de los alumnos,

alejado de sus intereses y experiencia.

El impacto de la evaluación sobre el contenido curricular y el enfoque de la enseñanza no debe considerarse necesariamente como negativo. Un sistema de evaluación eficaz apoya el aprendizaje en una variedad de formas, desde la retroalimentación formativa para el uso en decisiones de corto plazo acerca de las actividades de aprendizaje, hasta la entrega de información sobre los logros de los estudiantes para los padres y para su uso como evidencia en la planificación a largo plazo y en la auto-evaluación de la escuela. La evaluación también puede ayudar a clarificar y comunicar el sentido de los objetivos de aprendizaje, a través del establecimiento de criterios de logro o de la generación de tareas que ejemplifiquen el uso de habilidades de indagación y la comprensión.

Los impactos negativos surgen cuando lo que se evalúa refleja sólo los aspectos que pueden evaluarse fácilmente, agravado por la aplicación de premios y castigos a los resultados. Cuando los resultados tienen "alto impacto" para los profesores, esto pone presión sobre ellos, lo que se transfiere a los estudiantes, aunque las pruebas no sean de alto impacto para los estudiantes. Las investigaciones muestran que cuando esto sucede, los profesores centran la enseñanza en el contenido del examen, entrenan a los estudiantes para pasar los tests y se sienten impulsados a adoptar estilos de enseñanza que no coinciden con lo que se necesita para desarrollar una real comprensión. En la actualidad existe un gran cuerpo de evidencia científica sobre el impacto negativo del uso, con "alto impacto", de datos que provienen de la evaluación y las pruebas. El Recuadro 7 da una breve indicación de los resultados de una extensa revisión de la investigación sobre el impacto de las pruebas en los docentes y los alumnos.³⁵

Estos hallazgos plantean interrogantes acerca de la equidad, ya que las respuestas negativas a las pruebas no se distribuyen uniformemente en todos los estudiantes. Los estudiantes pueden estar en mayor desventaja por motivos de género, idioma, contexto familiar y habilidad general.

Un estudio a gran escala de la educación primaria en Inglaterra, llevado a cabo entre 2007 y 2009, que extrajo evidencia de una variedad de fuentes concluyó que las pruebas nacionales de los estudiantes al final de la escuela primaria (edad 11):

- ponen a los niños y profesores bajo una presión intolerable

35 Nordenbo, S. E., Allerup, P., Andersen, H. L., Dolin, J., Korp, H., Larsen, M. S., et al. (2009). *Pædagogisk brug af test - Et systematisk review*. København: Aarhus Universitetsforlag. (In English: Pedagogical use of tests – A systematic review). <http://www.dpu.dk/omdpu/danskclearinghouseforuddannelsesforskning/udgivelser/paedagogiskbrugaftest/>

- son altamente estresantes
- restringen el plan de estudios
- alteran los objetivos de aprendizaje
- socavan la autoestima de los niños
- son contrarias al compromiso de las escuelas con una educación completa y equilibrada
- convierten el último año de la escuela primaria en un año de saturación y pruebas.³⁶

Efectos similares se registraron en otras jurisdicciones donde los resultados de las pruebas se utilizan para hacer juicios sobre las escuelas y los profesores. Sin embargo, la investigación llevada a cabo en dos estados de Alemania, Hesse y Bremen, sugiere que el estrechamiento del currículo y la enseñanza en función de las pruebas, no resulta sólo de las evaluaciones de alto impacto, sino que acompaña la existencia misma de las pruebas. El estudio de Jäger et al. Sobre el efecto de las pruebas estatales de bajo impacto al final de la educación secundaria reveló que los maestros estaban usando la retroalimentación de las pruebas para mejorar los resultados de sus alumnos mediante la reducción de la amplitud curricular. Los investigadores sugieren que "Esto puede ser visto como una indicación de que los profesores se sienten responsables por los logros de sus estudiantes y por alcanzar las metas de logro y los estándares curriculares sin incentivos externos".³⁷

Las consecuencias para la ciencia son especialmente serias. Se ha observado que los profesores han adoptado un estilo de transmisión de la enseñanza, aunque se piense que esto no es lo mejor para ayudar a los estudiantes en la comprensión y el desarrollo de habilidades.³⁸

Otro de los impactos más serios reportados es sobre la práctica de la evaluación formativa. Después de la introducción de un régimen de pruebas en Inglaterra, los investigadores informaron que los estudiantes y los profesores estaban conscientes de que la evaluación en la sala de clases, en lugar de ser esencialmente formativa en su función, se convirtió en una serie de mini-evaluaciones sumativas.³⁹ Los profesores comprobaban el "nivel" de desempeño mucho más a menudo de lo necesario para la presentación de informes. Se desarrolló una "cultura orientada al logro", que deja poco margen para el uso de la evaluación formativa. Esta cultura fue considerada como responsable de que las escuelas voluntariamente sobrepasaran lo que se requiere en cuanto a la evaluación externa y los juicios acerca de los niveles alcanzados por los estudiantes.

La constatación de que la evaluación sumativa puede tener influencia sobre la evaluación formativa indica que prestar atención únicamente a la evaluación formativa, tendrá probablemente poco efecto. De hecho, la experiencia de los obstáculos para la introducción de la evaluación formativa genuina en los países donde existe una fuerte dependencia de las pruebas externas de alto impacto, es una demostración de esto. Así, si se busca mejorar el aprendizaje de la ciencia a través de ECBI y el uso de la evaluación formativa, es necesario también garantizar que la evaluación sumativa sea coherente con los objetivos de aprendizaje de ECBI.

36 Alexander, R. (Ed) (2010) *Children, their World, their Education*. Final report and recommendations of the Cambridge Primary Review. London: Routledge. P 316

37 Jager, J.J., Merki, K.M., Oerke, B. and Holmeier, M. (2012) State-wide low-stakes tests and a teaching to the test effect? An analysis of teacher survey data from two German States, *Assessment in Education*, 19 (4) 451-467, p 464

38 Osborne, J., Simon, S. and Collins, S.(2003) Attitudes towards science: a review of the literature and its implications, *International Journal of Science Education*, 25, 1049-1079

39 Pollard, A and Triggs, P. (2000) *Policy, Practice and Pupil Experience*. London: Continuum International

Box 7: Revisión sistemática de la investigación sobre el impacto de la evaluación de gran escala³⁵

La “Danish Clearinghouse for Educational Research” realizó una revisión sistemática sobre la investigación que aborda las siguientes preguntas:

1. ¿Cómo afecta la realización de pruebas a la enseñanza?
2. ¿Cómo se usan estos tests para propósitos pedagógicos por parte de los profesores?
3. ¿Cómo afecta a los estudiantes la realización de pruebas?

La revisión se basó en investigaciones publicadas en el período 1980-2008 y utilizó una metodología sistemática. Los principales hallazgos revelaron efectos negativos considerables sobre la enseñanza a causa de la introducción de las pruebas administradas centralmente:

- El currículum experimentado por los alumnos se reduce o distorsiona: los profesores simplifican sus exigencias respecto al pensamiento de los estudiantes; se enfatizan los hechos y las habilidades mecánicas en detrimento de las actividades creativas y estéticas
- Más tiempo de enseñanza se asigna a materias incluidas en los tests a expensas de aquellas no incluidas
- La enseñanza se dedica a enseñar para el test y al aprendizaje memorístico

El estudio también concluyó que, en general, los profesores no utilizan las pruebas a gran escala o los datos de las pruebas a gran escala en su enseñanza. De hecho, ellos miran estas pruebas con escepticismo y tienen actitudes negativas, a menos que hayan estado involucrados en su creación. La propia existencia de pruebas a gran escala puede reducir la disposición de los profesores a hacer uso de datos de este tipo de pruebas. Sin embargo, si los profesores sienten una cierta propiedad sobre la prueba estarán más dispuestos a utilizar los datos.

La influencia de los tests en los alumnos es dramática:

- El mero anuncio de un test genera respuestas emocionales como nerviosismo y miedo, especialmente entre las niñas
- Los alumnos se preparan para el test aprendiendo de memoria
- Para los alumnos de mayor desempeño aumenta la motivación, mientras que los de menor desempeño pierden motivación
- El resultado de un estudiante en la prueba puede influir en la motivación y la autoeficacia en el futuro.

El mito de elevar estándares a través de la realización de pruebas

Teniendo en cuenta estos efectos negativos, nos podemos preguntar: ¿por qué las pruebas nacionales tienen un papel tan importante en las políticas educativas para tanta gente? La respuesta está en la afirmación de que “las pruebas elevan los estándares”. Sin embargo, existe poca evidencia para apoyar esto y, de hecho, mucha que sugiere que el aumento en el puntaje de las pruebas se

debe a la familiaridad con la toma de la prueba y el enseñar para la prueba.⁴⁰ Los resultados de los exámenes pueden subir – al menos al principio – pero esto no da información sobre el cambio en el aprendizaje real. La consecuencia de focalizarse en lo que se evalúa, de entrenar para la prueba y de evaluar un rango restringido de contenidos, es que en realidad no es posible determinar, a partir de los resultados de pruebas nacionales, si los estándares nacionales han cambiado de año en año.

Este fenómeno, por el cual una medida a la que se le asigna "alto impacto" disminuye su valor como un indicador de la calidad de un servicio, es común a través de diversos servicios sociales. Se resume en la Ley de Goodhart,⁴¹ que suele enunciarse en forma sucinta como:

Cuando una medida se convierte en un objetivo blanco, deja de ser una buena medida.

En otras palabras, una vez que un indicador social o económico, u otra medida alternativa, se vuelve una meta como parte de la política, pierde la capacidad de dar información válida para su propósito original. Hay varios ejemplos en otros ámbitos distintos de la educación (por ejemplo, en el servicio de salud y los servicios de policía) donde un solo resultado fácilmente medible (por ejemplo, el tiempo de espera para recibir tratamiento en un hospital) se usa para evaluar un proceso complejo.

En el caso de educación, las pruebas son a menudo el medio elegido para fijar objetivos y el alto impacto conduce a prácticas que elevan los resultados de las pruebas sin una verdadera mejora en el aprendizaje, por lo tanto, la prueba deja de ser una medida útil de aprendizaje. Aunque más comúnmente ilustrado en términos de pruebas, el mismo efecto puede verse en la evaluación cuando el desempeño se juzga de acuerdo con los criterios. El uso de alto impacto de los resultados conduce a reducir la interpretación de los criterios que, a continuación, dejan de indicar la gama completa de competencias previstas.

Rendición de cuentas

Los efectos de juzgar a los profesores y a las escuelas sobre la base de los resultados de las pruebas son ampliamente reconocidos en la actualidad y dan lugar a los reclamos de que "la rendición de cuentas socava el aprendizaje". Sin embargo, esto es un error ya que el problema no es la rendición de cuentas, sino cómo ésta se ejerce. Cuando los resultados de las pruebas se toman como una medida de la calidad de la enseñanza, ello inevitablemente tendrá consecuencias para aquellos que están siendo juzgados. Puede pensarse que algunas de estas consecuencias no son intencionadas, como cuando los profesores interpretan demasiado estrechamente lo que se evalúa como algo importante de enseñar. Los responsables de los procedimientos de rendición de cuentas pueden alegar que esto no era lo que se intentaba, que "los profesores no necesitan enseñar para la prueba" y estar de acuerdo en que algunos de los impactos del régimen de pruebas no son deseables. Sin embargo, la práctica continúa. Uno puede sospechar que hacer responsables a los profesores por el aumento de los puntajes de las pruebas es un uso deliberado del impacto que tiene la evaluación en los contenidos curriculares; las pruebas proporcionan una forma directa de controlar lo que se enseña. Como comenta Stobart,⁴² esta es una forma más rápida y más barata de cambiar el contenido del currículo y de la pedagogía, que desarrollar el currículo y proporcionar desarrollo profesional. El mecanismo de esta influencia es establecer metas de elevación de resultados para las escuelas, los distritos o la nación en su conjunto.

A nivel del sistema, la ambición de muchos países de mejorar su posición en la "tabla de clasificación" de los resultados de PISA ha dado lugar a acciones que transfieren a los distritos y a las escuelas el

40 See, for example: Linn, R. L. (2000) Assessments and accountability, *Educational Researcher*, 29 (2) 4-16 and Tymms, P. (2004) Are standards rising in English primary schools? *British Educational Research Journal*, 30 (4) 477-94

41 <http://www.atm.damtp.cam.ac.uk/mcintyre/papers/LHCE/goodhart.html>

42 Stobart, G. (2008) op cit p138

imperativo de cambiar las prácticas.⁴³ En el ámbito nacional, el uso de pruebas parece tener el efecto deseado durante unos pocos años. Las pruebas nacionales para los estudiantes al final de la escuela primaria en Inglaterra son un ejemplo. En 1995 se introdujo ensayos en inglés, matemáticas y ciencias por cada estudiante de 11 años. Durante los primeros cinco años, los puntajes aumentaron año tras año, pero a partir de 2000 no hubo ningún cambio apreciable. Un estudio exhaustivo⁴⁴ de otros datos sobre los cambios en el desempeño a lo largo de estos años (incluidos los datos de las encuestas TIMSS) no mostró mejoramiento en el desempeño. La conclusión alcanzada fue que las razones más probables para los cambios iniciales fueron el efecto de enseñar la técnica para enfrentar la prueba (las pruebas formales eran nuevas para los alumnos de esta edad en 1995) y focalizar la enseñanza en el contenido de la prueba.

Se han reportado hallazgos y conclusiones similares en otros regímenes de pruebas, por ejemplo en EE.UU. Linn encontró "un patrón de ganancias tempranas seguidas por una estabilización"⁴⁵ como comportamiento típico a lo largo de todos los Estados en los cuales se utilizaron las pruebas de alto impacto. Stobart describe el proceso de estabilización, a través del cual el efecto se degrada, refiriéndose a la "vida media" de la examinación para la rendición de cuentas. Esta sería, probablemente, alrededor de cuatro años, durante los cuales:

La laxitud se ha eliminado; los profesores se han familiarizado con la prueba; y los estudiantes saben de exámenes de años anteriores lo que tienen que hacer, por lo que el aprendizaje se reduce cada vez más a la forma de maximizar las notas.⁴⁶

En la escuela, la obligación de cumplir con los objetivos en forma de los resultados de pruebas significa que mayor atención se presta a aquellas materias sometidas a las pruebas respecto a las otras. Para las escuelas primarias, donde la lectura, la escritura y las matemáticas son invariablemente blanco de las pruebas nacionales, esto tiene serias consecuencias para la ciencia y otros temas. De hecho, cuando en Inglaterra se terminaron las pruebas al final de la escuela primaria en 2010 - por motivos relacionados con la escasa validez de las pruebas, que eran escritas y no requerían ninguna experiencia práctica - el efecto fue la transferencia de horas de clase y atención a inglés y matemáticas, que sigue siendo objeto de ensayos de alto impacto.

Lo que es importante evitar es el uso inapropiado de los resultados de las pruebas de los estudiantes, como cuando éstas se toman como los únicos indicadores en la evaluación de los profesores y de las escuelas. La razón para esto es simplemente que lo que los estudiantes logran no sólo es el resultado de sus experiencias en la escuela.

Los límites de la rendición de cuentas

La rendición de cuentas significa que se es responsable de las propias acciones y que se es capaz de explicar a los interesados por qué y cómo se hicieron ciertas cosas o por qué no se hicieron. Los profesores sólo pueden ser considerados responsables de las acciones o de los resultados en aquello que tienen control, como lo que hacen en el aula, las oportunidades de aprendizaje que ofrecen y la ayuda que dan a los estudiantes, y así sucesivamente. No son necesariamente responsables sobre si se alcanzan aprendizajes prescritos externamente, ya que esto depende de muchos otros factores, sobre los que el profesor no tienen control, como el aprendizaje previo de los estudiantes y las

43 Ertl, H.(2006) Educational standards and the changing discourse on education: The reception and consequences of the PISA study in Germany. *Oxford Review of Education*, 32 (5) pp 619–634

44 Tymms, P. (2004) Are standards rising in English primary schools? *British Educational Research Journal*, 30 (4):477–494.

45 Linn, R.L. (2000) Assessment and accountability, *Educational Researcher*, 29 (2), 4–16, p6

46 Stobart, G. (2008) op cit p134

muchas influencias y condiciones externas a la escuela que afectan a su aprendizaje. Estos factores deben ser tenidos en cuenta tanto por los profesores en su planificación y su trabajo para alcanzar las metas de aprendizaje de los alumnos, y por aquellos que hacen a los profesores responsables de la calidad de la educación de los estudiantes. Se desprende de esta argumentación que la información utilizada en la rendición de cuentas debe incluir, además de los datos sobre los logros de los estudiantes, información sobre el currículum y los métodos de enseñanza, y aspectos relevantes sobre los contextos de los alumnos y sus historias de aprendizaje⁴⁷.

Recuadro 8: Las escuelas haciendo su propio relato

Una aproximación positiva a la rendición de cuentas permite a las escuelas y a los profesores expresar sus propias metas y sus logros en relación con ellas. Se espera actualmente que las escuelas comuniquen y expliquen su filosofía, los objetivos y las políticas a la comunidad en general. Esta es una oportunidad para que el logro de los estudiantes se presente en el contexto de los muchos factores que afectan a este logro. El proceso se ve facilitado en algunos países por orientaciones⁴⁷ para las escuelas, lo que les da un marco para contextualizar su propio relato, diseñado para fomentar la auto-evaluación de las escuelas y un papel más formativo para todo el proceso. En algunos casos, esto es monitoreado por inspectores, en otros está disponible para el escrutinio público.

Los datos sobre los logros de los estudiantes, derivados de los registros de los docentes y de la evaluación en relación con una amplia gama de objetivos de aprendizaje, no sólo de los resultados de las pruebas, serán parte de la cuenta que dé la escuela sobre su desempeño a los padres y otros que tengan un interés en esto. La cuenta también debe mostrar cómo, en el marco de su evaluación interna, los datos agregados sobre los logros de los estudiantes se utilizan para comprobar el progreso de grupos de estudiantes, como los que tienen lenguas maternas diferentes, aquellos con tendencias a alcanzar mayores o menores logros, y los niños y las niñas. Todo esto depende de contar con medidas efectivas en relación con los objetivos de aprendizaje y posicionar los resultados en el contexto de las acciones y de los sistemas utilizados para lograr estos objetivos.

Evaluación y teorías del aprendizaje

La alineación de la evaluación con el contenido del currículum y de la pedagogía es importante cuando se considera la validez de la evaluación. En la discusión en el Capítulo 1 se indica que la validez es un juicio de la medida en que las inferencias que se hacen están apoyadas por la evidencia y fundamentos teóricos. Un punto que se expresó con frecuencia en la conferencia (véase la Introducción) fue la necesidad de contar con una evaluación que reflejara la teoría del aprendizaje que subyace a ECBI. Necesitamos, por lo tanto, incorporar las teorías sobre el aprendizaje a la discusión sobre el impacto de la evaluación en el contenido del currículum y la pedagogía.

47 SEED (Scottish Executive Education Department) (2002) How Good is Our School? Self evaluation using quality indicators. Edinburgh: HMIE.

Teorías del aprendizaje

Las muchas y variadas teorías del aprendizaje pueden agruparse en tres tipos: "conductista", "cognitivo constructivista" y "socio-cultural constructivista". (En alguna literatura de EE.UU.,⁴⁸ constructivista se describe como "cognitiva" y socio-cultural como "situada", pero los significados son básicamente los mismos.) Una simple formulación de estos términos, basada en Watkins,⁴⁹ expresa su significado como:

- Conductismo: "Aprender es recibir enseñanza"
- Constructivismo cognitivo: "Aprender es dar sentido en forma individual "
- Constructivismo socio-cultural: "Aprender es construir conocimiento como parte de hacer cosas con otros."

El *conductismo* describe una visión del aprendizaje en el que los comportamientos se forman por un sistema de premios y castigos. El supuesto es que los comportamientos que se premian serán reforzados y aquellos que se castigan desaparecerán, el aprendizaje puede controlarse externamente y la motivación es casi en su totalidad extrínseca. Una característica adicional para el conductismo, que es particularmente relevante para la evaluación, es que los comportamientos complejos son deconstruidos en partes que se pueden enseñar, practicar y evaluar por separado. Este punto de vista, entonces, es consistente con pruebas de hechos y habilidades inconexas, donde la velocidad es parte de la esencia y las respuestas son correctas o incorrectas.

Dos características claves de la perspectiva *cognitivo constructivista* del aprendizaje son que los aprendices construyen su propio conocimiento mediante el desarrollo de modelos mentales y que el conocimiento existente tiene un papel importante en este desarrollo. El objetivo es la comprensión, que es vista como algo que ocurre cuando la nueva experiencia se incorpora en un modelo existente o nuevo. La participación activa de los estudiantes se considera primordial, ya que, como se cita frecuentemente, "ellos hacen el aprendizaje". Las visiones constructivistas del aprendizaje sustentan la evaluación formativa, que parte de "averiguar donde los estudiantes están en su aprendizaje" con el fin de decidir "a donde tienen que ir y la mejor manera de llegar allí "(ver página 20). Hay pocos ejemplos de la evaluación sumativa que se basen en una visión constructivista del aprendizaje, aunque existen algunos intentos a través de pruebas computacionales adaptativas y mapas conceptuales en formato computacional (*screen-based concept-mapping*)⁵⁰. James concluye que "la mayor parte de las pruebas formales descansa todavía en gran medida en enfoques conductistas".⁵¹

En la perspectiva *constructivista socio-cultural* sobre el aprendizaje también hay un foco en la comprensión, pero a través de "dar sentido a la nueva experiencia con los demás, en lugar de a través del trabajo individual". En estas situaciones, el individuo toma de (internaliza) una experiencia compartida lo que se necesita para ayudar a su comprensión y a continuación exterioriza el resultado como un aporte a la discusión en grupo. Hay un constante ir y venir desde el individuo al grupo y el conocimiento se construye en comunidad a través de la interacción social y el diálogo. Los recursos físicos y el lenguaje también tienen roles importantes, como James explica:

De acuerdo con esta perspectiva, el aprendizaje se produce en las interacciones entre el individuo y el entorno social. El pensamiento se lleva a cabo a través de acciones que alteran la situación y la situación cambia el pensamiento; los dos interactúan

48 Discussed in Pellegrino, J.W., Chudowsky, N. and Glaser, R. (Eds) (2001) *Knowing what Students Know - The Science and Design and Educational Assessment*. Washington, DC: National Academy Press.

49 Watkins, C. (2003) *Learning: A Sense-Maker's Guide*. London: Association of Teachers and Lecturers.

50 Osmundson, E., Chung, G., Herl, H., Klein D. (1999) Knowledge-mapping in the classroom: a tool for examining the development of students' conceptual understandings. Los Angeles, California: National Centre for Research on Evaluation and Student Testing, University of California. www.cse.ucla.edu/Reports/TECH507.pdf

51 James, M. (2012) Assessment in harmony with our understanding of learning: problems and possibilities, in Ed J.Gardner *Assessment and Learning*, 2nd edn. London: Sage pp187 – 205.

constantemente. Especialmente importante es la noción que el aprendizaje es una *actividad mediada* en la que los productos culturales tienen un papel crucial. Estos pueden ser objetos físicos tales como libros y equipamiento, pero también pueden ser herramientas simbólicas tales como el lenguaje. Puesto que el lenguaje, que es fundamental para nuestra capacidad de pensar, se desarrolla en relaciones entre las personas, las relaciones sociales son necesarias para, y preceden, el aprendizaje (Vygotsky, 1978). Así, el aprendizaje es una actividad social y de colaboración en la que las personas desarrollan su pensamiento en conjunto.⁵²

De la perspectiva de Vygotsky⁵³ se desprenden algunas implicancias profundas para la evaluación, en cuanto para cualquier aprendiz hay un área justo más allá de su comprensión actual (donde una persona está en control consciente de las ideas y sabe que él o ella las está usando) en la cual ideas más avanzadas pueden ser utilizadas con ayuda. Vygotsky llamó a esta zona "la zona de desarrollo próximo (o potencial)". Es, en esencia, lo que hemos llamado el "siguiente paso" que se espera dé el estudiante y que es identificado a través de la evaluación formativa. "Andamiaje" es un término muy adecuado que se usa para describir cómo ayudar a los estudiantes a dar el siguiente paso en la comprensión a través de la introducción de ideas nuevas o mejores prácticas científicas, proporcionando a los estudiantes vocabulario que les permita expresar sus ideas con mayor precisión.

El reconocimiento que, en compañía de otros aprendices, los estudiantes pueden exceder lo que pueden entender y hacer por sí mismos, genera dudas respecto a cuál es su "verdadero" nivel de desempeño ¿es el nivel de "desempeño independiente" o el nivel de "desempeño asistido" en el contexto social? Se ha argumentado que el nivel de desempeño estimado cuando se responde a la asistencia y a las nuevas herramientas proporcionadas por otros entrega una mejor evaluación que la administración de tests no asistidos.⁵⁴

Recuadro 9: Un modelo de aprendizaje de las ciencias a través de la indagación

En el aprendizaje basado en la indagación el desarrollo de la comprensión deriva de la curiosidad sobre un fenómeno o evento que es nuevo para los alumnos y que plantea preguntas que captan su atención. La exploración inicial puede revelar situaciones que traen a la mente una idea de la experiencia previa que sugiere una posible explicación o una respuesta a una pregunta. Puede ser la idea de un individuo en particular o el resultado de la lluvia de ideas con otros estudiantes o de fuentes de consulta de información. Trabajar científicamente involucra hacer una predicción basada en la idea y luego reunir datos relevantes para ver si hay evidencia para apoyar la predicción y la aplicación de la idea. Esto podría ser una larga investigación que implique la experimentación controlada o una simple extensión de las observaciones.

Encontrar que la evidencia calza con la predicción y que la idea proporciona una buena explicación significa que esta idea se ha convertido en "más grande", ya que entonces explica una gama más amplia los fenómenos. Incluso cuando parece "que no funciona", algo se ha aprendido acerca de su rango de aplicación. Pero para encontrar una explicación que "funcione" se deben someter a prueba ideas alternativas. Esto puede venir de la lluvia de ideas inicial o posterior informada por lo que se ha encontrado. La utilidad de las ideas desarrolladas de esta forma depende de la recopilación y el uso de la evidencia de una manera científica. Así, la capacidad para usar habilidades científicas de indagación es una parte esencial del desarrollo de la comprensión y el resultado del pensamiento compartido sobre qué datos recopilar y cómo hacer para recopilarlos e interpretarlos.

52 ibid p 192/3

53 Vygotsky, L.S. (1978) *Mind in Society: The Development of Higher Psychological Process*. Cambridge, MA: Harvard University Process

54 Grigorenko, E. (1998) Mastering tools of the mind in school (trying out Vygotsky's ideas in classrooms), in Eds R. Sternberg and W. Williams *Intelligence, Instruction and Assessment: Theory and Practice*. Mahwah, NJ: Erlbaum.

Implicancias de ECBI para la evaluación

En el Recuadro 9, ECBI se describe como un proceso de construcción de comprensión a través de la recopilación de evidencia para someter a prueba posibles explicaciones y las ideas subyacentes, y que tiene elementos de las perspectivas constructivistas y socio-culturales del aprendizaje. Si la validez depende de cuán bien la evaluación refleja la visión del aprendizaje implícito en ECBI, ¿cómo debería llevarse a cabo la evaluación sumativa del aprendizaje basado en la indagación? Evidentemente no es con estudiantes sentados en forma aislada unos de otros en una sala de exámenes. Una evaluación coherente con la visión del aprendizaje que sirve de fundamento a ECBI debería basarse en lo que los estudiantes pueden hacer en la interacción con los demás. Cuando las actividades en el aula o en el laboratorio involucran andamiaje por parte de los profesores e interacciones entre los estudiantes, estas actividades ofrecen oportunidades para evaluar el “desempeño asistido” de los estudiantes y añaden una razón más en favor de que los profesores tengan un papel central en la evaluación.

Un ejemplo de la evaluación informada por la interacción social y el uso de los artefactos es el enfoque utilizado en la investigación de Dolin y Krogh, la que involucró la re-examinación de estudiantes en algunos ítems de PISA (ver página 70). El estudio incluyó una entrevista / conversación con cada estudiante por unos 30 minutos, para explorar su conocimiento y razonamiento. Sin embargo, el uso de estas situaciones uno a uno es claramente de aplicación limitada cuando se evalúa un gran número de estudiantes, y presenta problemas de fiabilidad. Una situación más controlada sería que pares o grupos de estudiantes recibieran una tarea y algo de tiempo para una lluvia de ideas acerca de cómo abordarla y, si fuese posible, para llevar a cabo una investigación. Cada estudiante produciría a continuación un reporte individual del trabajo que se evaluaría usando una serie de criterios

Un enfoque diferente puede ser que la "interacción" tenga lugar a distancia. Los estudiantes realizan una investigación y reciben comentarios sobre esta investigación. Las respuestas de los estudiantes a los comentarios contribuyen a la evaluación general. Estos planteamientos, de los cuales hay pocos ejemplos en la práctica, desafían las visiones convencionales sobre el proceso y el significado de la evaluación. Como hace ver James “Todavía queda mucho trabajo por hacer para encontrar la manera de llevar la evaluación a una mejor alineación con algunas de las ideas más poderosas de la teoría contemporánea del aprendizaje”.⁵⁵ En el Capítulo 7 se discute lo que parece ser necesario para que el cambio ocurra.

El rol de la evaluación en la comprensión de los principales problemas globales

Afirmar que la evaluación puede ayudar al desarrollo de la comprensión de los principales problemas globales, tales como el calentamiento global, la pérdida de la diversidad de organismos, el hambre y la pobreza causada por la actividad humana, puede parecer exagerando el impacto de la evaluación. Sin embargo, la educación tiene sin duda un papel clave en la generación de los cambios necesarios para hacer frente a estos problemas. Por lo tanto, la evaluación, que tiene un papel clave en la educación, también debe tener una parte. Un objetivo fundamental de la educación es preparar a los estudiantes para ser ciudadanos informados que entienden las razones de los problemas creados por la actividad humana y lo que se requiere para solucionarlos, y que están motivados para participar en la acción responsable. Hay muchos aspectos a tener en cuenta, pero dos preguntas importantes son:

- ¿Cuáles son las ideas esenciales o conceptos que necesitan ser comprendidos?

55 James, M. (2012) op cit p196

- ¿Cómo se puede generar la comprensión (en oposición al conocimiento superficial de los hechos relacionados) de estos conceptos esenciales o centrales?

Para responder la primera pregunta es necesario identificar las ideas que son relevantes y poderosas para ayudar a la comprensión del mundo y de cómo éste funciona, cómo interactúan sus componentes, cómo la intervención humana puede y no puede influir en nuestro entorno global. Esto significa la identificación de las "grandes" ideas *de* la ciencia y *sobre* la ciencia (es decir, cómo funciona la ciencia, sus fortalezas y limitaciones) y garantizar que la educación en ciencias está diseñada para desarrollar la comprensión de estas ideas.

La mejor respuesta que tenemos para la segunda pregunta, de cómo lograr la verdadera comprensión, es que los estudiantes aprenden a través de la indagación. Hemos descrito en el Capítulo 2 y en el Recuadro 9, el significado del aprendizaje a través de la indagación, argumentando que permite el aprendizaje activo (física y mentalmente) a través del cual los estudiantes dan sentido a sus investigaciones sobre el mundo que los rodea. Esto no significa aprender por sí mismos, o tener que resolverlo todo por sí mismos - ni mucho menos - porque, como se señaló anteriormente, la visión actual de cómo las ideas se desarrollan apuntan al papel de la colaboración, la discusión y el diálogo con otros, en el que las ideas se avanzan por el pensamiento colectivo y basándose en lo que ya se ha encontrado en las actividades de los científicos durante generaciones. Lo que sí significa es que cada alumno está desarrollando una comprensión compartida que le hace más sentido que otros puntos de vista alternativos, los cuales ellos (u otros) previamente pudieron haber sostenido.

Entonces, ¿en qué momento aparece la evaluación en este argumento? En breve, si el principio de que "todas las evaluaciones en última instancia deberían ayudar el aprendizaje" (véase el Recuadro 21 en el Capítulo 7) se traduce a la práctica, su rol se vuelve transparente. Más detalladamente, utilizar la evaluación formativa regula la enseñanza y el aprendizaje, de modo que la comprensión se ve apoyada. Esta actúa para optimizar el desafío de las nuevas experiencias de los estudiantes a fin de que no estén ni demasiado lejos de sus ideas y competencias existentes, ni sean demasiado familiares de modo que haya poco cambio en lo que pueden hacer o saber.

En relación con la evaluación sumativa hay muchas formas en las que ésta puede ser utilizada para apoyar el aprendizaje. Por ejemplo, como se mencionó en el Capítulo 3, mapear el progreso individual de los estudiantes y de grupos de estudiantes alerta a todos los involucrados en cuanto a la necesidad de adoptar medidas para asegurar oportunidades de aprendizaje para todos, para las niñas y los niños, para los más capaces y los menos capaces, para los desfavorecidos, así como para los más afortunados. Lo que se mapee debe reflejar el objetivo de garantizar que todos los estudiantes están haciendo progresos hacia las ideas más importantes y las habilidades de indagación científica que son parte de las razones para promover ECBI. (Más adelante, en el Capítulo 6, observamos algunas implicancias de adoptar este objetivo en la forma de reportar los logros de los estudiantes). El progreso de los estudiantes hacia estos objetivos es un factor clave en la auto-evaluación de las escuelas en cuanto a su aporte para que los alumnos aprendan con comprensión.

La evaluación es sólo uno de los diversos factores que influyen en el aprendizaje de los estudiantes. Es evidente que lo que los estudiantes pueden alcanzar depende del contenido del currículo y la pedagogía, pero las interacciones que se indican en la Figura 3 y la discusión anterior en este capítulo muestran que la evaluación no puede ser ignorada. Por otra parte tiene algo que aportar a una mejor comprensión de los objetivos, a través de la expresión de lo que significa alcanzarlos. La evaluación, entonces, tiene que ser parte de la discusión sobre la manera de proporcionar educación de relevancia para enfrentar los problemas globales.

Capítulo 5

Implementando la evaluación formativa de ECBI

El Capítulo 3 considera, en términos generales, el significado y la importancia de la evaluación *para* el aprendizaje (formativa) y *de* aprendizaje (sumativa). En este capítulo y en el siguiente nos preocupamos de los métodos utilizados para llevar a cabo la evaluación para estos fines, pero específicamente en el contexto de la evaluación de los objetivos de ECBI. En el Capítulo 2 se identificaron seis objetivos de la educación en ciencias como esenciales para preparar a los estudiantes para la vida en un mundo rápidamente cambiante como el actual, mientras que la definición de ECBI en la página 13 destaca su contribución específica al desarrollo de la *comprensión* y de las *habilidades* usadas por los *científicos* (que llamaremos *habilidades indagatorias de la ciencia*). Así, el foco de este capítulo es sobre cómo la evaluación puede apoyar el desarrollo en función de estos dos objetivos: la comprensión científica y las habilidades indagatorias de la ciencia. El Capítulo 6 se centra en los métodos de evaluación sumativa y en la forma de reportar el progreso hacia estas metas

Evaluación formativa y ECBI

La implementación de la evaluación formativa no es fácil. No es una cuestión de utilizar determinados materiales o técnicas, sino que requiere habilidades y conocimientos por parte de los profesores. El propósito de la evaluación formativa es identificar y ayudar a los estudiantes a dar, los próximos pasos en el progreso de desarrollar su comprensión y competencia. Para implementar la evaluación formativa, los profesores necesitan saber cómo obtener y utilizar la evidencia sobre el progreso de los estudiantes en el aprendizaje y cómo proporcionar ambientes de aprendizaje efectivos que apoyen nuevos progresos. Algunos investigadores, siguiendo a Sadler,⁵⁶ describen esto como tomar medidas para que el estudiante pueda "cerrar la brecha" entre el estado actual de comprensión y competencia, y la meta de aprendizaje. Sin embargo, referirse a los "próximos pasos" en lugar de "cerrar la brecha" es una mejor manera de transmitir una visión sobre el progreso en el aprendizaje como un proceso continuo. Además, y esto es crucial, la evaluación formativa requiere un cambio en cómo los profesores ven el proceso de aprendizaje y su rol en él. Cuando el aprendizaje es visto como algo que *hacen* los estudiantes, no algo que se *les hace* a ellos, el papel del profesor o la profesora es diseñar ambientes en los cuales los estudiantes puedan participar activamente en construir su comprensión y desarrollar las competencias.

Es fácil darse cuenta que la evaluación formativa es esencial para la implementación de ECBI. Aprender a través de la indagación es un proceso de desarrollo de comprensión que toma en cuenta la forma en que los estudiantes aprenden mejor, es decir, a través de su propia actividad física y mental. Se basa en el reconocimiento que las ideas, el conocimiento y la comprensión son contruidos por los estudiantes a través de su propio pensamiento acerca de sus experiencias. Lo que se sabe acerca del aprendizaje nos dice que esto sucede cuando las actividades permiten que los estudiantes desarrollen su comprensión, es decir, cuando están trabajando en el área entre las ideas existentes y las ideas y competencias más avanzadas, o en la zona de desarrollo potencial (véase la

56 Sadler, D. R. (1989) Formative assessment and the design of instructional systems, *Instructional Science*, 18, 119-44.

página 37). La evaluación formativa es la estrategia mediante la cual estas actividades y todo el ambiente de aprendizaje están diseñados para asegurar que este progreso en el aprendizaje sea posible. Cuando se aprende a través de la indagación los estudiantes desarrollan su comprensión a través de las actividades indicadas en el Recuadro 3 y en la definición de ECBI: hacer observaciones, plantear preguntas investigables, planificar y conducir investigaciones, revisar la evidencia a la luz de lo ya conocido, sacar conclusiones, comunicar, y sostener discusiones en las que las ideas se comparten, se explican y defienden. También se reconoce que la participación en estas actividades depende de la medida en la cual generan interés, son percibidas como relevantes, se disfrutan y entusiasman.

Para el profesor, la evaluación formativa en ECBI significa utilizar las estrategias identificadas en la descripción en el Capítulo 3:

- promover el diálogo en el aula
- utilizar preguntas para generar evidencia sobre (y para ayudar al desarrollo de) las ideas y las competencias de los estudiantes
- proporcionar retroalimentación a los estudiantes
- usar la retroalimentación a los estudiantes para regular la enseñanza
- animar a los estudiantes a participar en la evaluación de la calidad de su trabajo

Aquí consideramos cómo estos aspectos de la evaluación formativa pueden ser implementados por los docentes. Más adelante, en el Capítulo 7, volvemos a la cuestión de cómo lograr los cambios que se necesitan en la enseñanza para implementar estas prácticas.

Diálogo en el aula

Es a través del lenguaje que desarrollamos una comprensión compartida de las ideas. Las ideas que podemos formar a partir de la experiencia directa tienen que ser comunicadas y esto involucra tratar de encontrar las palabras que expresen lo que queremos decir a los demás. En este proceso, nuestras propias ideas a menudo tienen que ser reformuladas al ser influidas por el significado que otros dan a las palabras.

La valoración por el aprendizaje de la conversación se estableció en la antigüedad, pero en los tiempos modernos fue la investigación de Douglas Barnes en la década de 1970, la que llamó la atención a la importancia de la conversación informal o "exploratoria".⁵⁷ En este tipo de conversaciones los estudiantes se interrumpen unos a otros, se repiten, vacilan y rephrasean. Barnes sugiere que los estudiantes sólo se involucran en este tipo de conversación en ausencia de la profesora o profesor porque entonces no hay ninguna fuente de autoridad a la que los alumnos pueden acudir. Sin embargo, también mostró que el profesor tiene un rol en garantizar que el pensamiento se ponga a prueba, la vaguedad se desafíe y se asegure que los estudiantes tengan una base para sus afirmaciones. El profesor, entonces, tiene que transitar un camino cuidadoso entre la dominación de la discusión, que desalentaría a los estudiantes de desarrollar y expresar sus propias ideas, y dejar a los estudiantes sin el estímulo para extender su pensamiento.

El extracto en el Recuadro 10 corresponde a la transcripción de la discusión de un grupo de tres niñas observadas por Barnes, antes y durante la intervención por parte del profesor. Barnes señala que en las preguntas "¿Qué presión del aire?" "¿Por qué no apareció antes?" "¿En qué momento se detuvo?" El profesor estaba estimulando a los jóvenes a considerar el proceso sistemáticamente, proporcionando preguntas que no se habían preguntado. Sin embargo, los intercambios iniciados

57 Barnes, D. (1976) *From Communication to Curriculum*. Harmondsworth: Penguin.

por el profesor ilustran un patrón común en el discurso de aula, en los que:

- el profesor hace una pregunta (¿qué presión del aire?)
- el estudiante responde (dentro de la botella)
- el profesor realiza un comentario evaluativo (bien..., entonces)
- hace otra pregunta (¿por qué no apareció antes?)

y el patrón de "pregunta-respuesta-evaluación" se repite. Este discurso implica mucha interacción entre el profesor y los estudiantes, pero en ella la autoridad del profesor es inconfundible. Contrasta esto con el papel del profesor en el Recuadro 11, donde el profesor también alienta un pensamiento más profundo, el uso de la evidencia y la claridad de significado, pero compartiendo la participación en los intercambios, en lugar de dominándolos. El profesor no evalúa las respuestas de los estudiantes, pero los lleva a expresarlas y explicarlas.

Recuadro 10: Explicando el efecto de la presión del aire

Después de algunas lecciones sobre el aire, se solicitó a los estudiantes (edad 12 y 13 años) que realizaran algunas actividades simples que involucraban la presión del aire. En una de éstas el equipamiento era una botella que contenía agua, cerrada por un tapón con una paja que lo atravesaba situada justo bajo la superficie del agua contenida en la botella. Se les pidió a los estudiantes (T, B y C) que soplaran fuertemente en la paja y que se alejaran para ver lo que ocurría.

Discusión con el profesor

- T Ah! El agua sube por la paja.
 C ¿Me pregunto cómo... me pregunto por qué... Cómo se para?
 ? ...va a detenerse
- Otra niña sopla en la paja*
- B Qui... quizás se debe a la presión del aire
 T Si... debe ser la presión del aire y...
 C ... es la presión del aire
 C está presionando en el agua hacia abajo

Discusión cuando la profesora se une al grupo

- B *(en respuesta a la profesora preguntando qué pasó)*
 T sopló hacia abajo y burbujeó hacia arriba y luego sacó su boca y todo subió por la presión del aire
- Profesora ¿Qué presión del aire?
 B La ... eeh... dentro de la botella
- Profesora Bien. Ahora, ¿por qué no había subido antes entonces?
 C y T porque no había suficiente aire... presión de aire
- Profesora ... No había suficiente aire dentro, pero cuando soplaron hacia dentro hubo más aire y lo forzó... ¿Por qué se detuvo? Ahora no está ocurriendo... ¿Por qué no está ocurriendo?

Alexander ha identificado el rol de la profesora en tales interacciones verbales, como "enseñanza dialógica". Él lo describe como "un enfoque pedagógico diferenciable", que "aprovecha el poder de

la conversación para estimular y extender el pensamiento de los niños, y para avanzar en su aprendizaje y comprensión⁵⁸. También permite al profesor diagnosticar y evaluar con mayor precisión. "Es a través de la enseñanza dialógica que los profesores pueden "dirigir la conversación en la clase con objetivos educacionales específicos en mente".⁵⁹ En relación con la ciencia esta "dirección" se centra en el uso de evidencia y puede conducir a lo que ha sido descrito como "argumentación". Esto es diferente de argumentar en la vida cotidiana:

En la ciencia, los objetivos de la argumentación son promover la mayor comprensión de una situación que sea posible y persuadir a los colegas de la validez de una idea específica. En lugar de tratar de ganar una discusión, como las personas hacen con frecuencia en contextos no científicos, la argumentación científica es, idealmente, sobre compartir, procesar y aprender acerca de las ideas.⁶⁰

El Recuadro 11 presenta un ejemplo de discurso dialógico entre un profesor y dos niñas (de 10 años) tratando de utilizar evidencia para distinguir entre los huevos crudos y cocidos. Tras acordar que C es el huevo crudo, D y M no están de acuerdo acerca de la identidad de los otros dos huevos. M tiene una razón para considerar que B es duro, sobre la base de que "los malos suben hasta arriba", por lo que considera que a B se le ha hecho algo. Pero ella no expresa las consecuencias de esto hasta que D trata de darle su razón. Entonces es como si la razón de D, que ella interrumpe, disparara su propio pensamiento.

N este caso el profesor hace poco, excepto alentar a las niñas en su esfuerzo por producir su respuesta y explicar su razonamiento. Sólo el ocasional "¿Por qué piensas eso?" o el reconocimiento "ya veo", y el refuerzo- "un poco crudo", alienta el uso de evidencias en apoyo de sus argumentos. Esto se observa con más claridad cuando M dice "si trabajamos en base al principio de que...", cuando ella relata lo que predice basándose en su juicio sobre la observación de la rapidez con que el huevo sube a la superficie en el agua salada, pero también ocurre a lo largo de la interacción. Vale la pena señalar, de paso, que el origen de su idea es el conocimiento previo acerca de cómo distinguir los huevos "buenos" de los "malos".

En la evaluación formativa, donde el objetivo es que los estudiantes revelen el desarrollo de la comprensión de un fenómeno o evento, la conversación entre los estudiantes es una fuente de información clave. El tipo de pensamiento exploratorio y el diálogo que Barnes y Alexander han abogado se estimula en un clima de aula en el que los profesores:

- esperan que los estudiantes expliquen las cosas
- valoran las ideas de los estudiantes aún si éstas no están bien formadas y son muy conjeturales
- evitan dar la impresión que sólo la respuesta "correcta" es aceptable y que los estudiantes deben adivinar
- juzgan cuándo intervenir o cuándo es mejor dejar que la discusión entre los estudiantes continúe sin interrupción.

58 Alexander, R. (2004) *Towards Dialogic Teaching. Rethinking Classroom Talk*. Cambridge: Dialogos, p1

59 *ibid*: p 27

60 Michaels, S., Shouse, A.W. and Schweingruber, H.A (2008) *Ready, Set, Science! Putting research to work in K-8 Science Classrooms*, Washington: National Academies Press, p89.

*Recuadro 11: Conversación dialógica sobre huevos*⁶¹

Dos niñas diez años de edad, D y M, estaban investigando la forma en que tres huevos de gallina enteros", denominados A, B y C se comportan en agua de la llave y agua salada. Sabían que uno era un huevo duro, uno "pasado por agua" (o semi-duro) y otro crudo. Tenían que averiguar cuál era cuál.

Cuando los huevos fueron colocados en el agua salada, el que estaba marcado como B flotó en la superficie, C se encontraba en el fondo y A suspendido a mitad de camino. La transcripción comienza cuando el profesor se acerca después de que han estado trabajando solos durante algún tiempo.

- Profesor ¿Me pueden decir cómo les está yendo?
- D Creo que C es el crudo
- M Los dos pensamos que C es el crudo
- Profesor ¿Eso piensan?
- D B es...
- Profesor (a D) ¿Por qué piensas eso?
- M Porque cuando se ponen huevos en agua los malos suben hasta arriba
- D (al mismo tiempo) Porque... los ponemos todos en ...
- Profesor ¿Malo?
- M Si, eso pienso— ¿o son los buenos? ... bueno, no sé
- Profesor ¿Si?
- M ... subieron hasta arriba, entonces...
- (D pone los huevos en agua salada)
- D ... este está al fondo (apuntando a C)
- M ... si está crudo debería quedarse al fondo
- Profesor Ya veo
- D Entonces eso es lo que pensamos, C está crudo y B está medio y A está duro
- M ... y yo creo que B está duro y ella cree que B es medio
- Profesor Ah, ya veo (a D) Puedes explicar, entonces, ¿por qué piensas eso?
- D Si ponemos... er... saca C (saca C, lo pone en la mesa, entonces levanta A y B) y los coloca adentro uno después del otro. Pon A adentro— no B primero. Esto es lo que ... Allyson piensa que está duro, yo pienso que está medio. Si lo pones en ... (ella pone B en el agua salada)
- M ...porque subió más rápido
- D Sube rápido. Y si pones éste adentro...
- (Ella pone A dentro del agua salada. Se va al fondo y sube muy lentamente)
- M Y este sube más lentamente
- D Por eso, yo creo que (mostrando a A) está duro porque ... bueno ...
- M Yo no. Pienso que si trabajamos en base al principio que éste (mostrando a B), entonces éste sube más rápido porque está, tú sabes, no verdaderamente hervido. Está un poco crudo
- Profesor Un poco crudo
- M Entonces, va a subir más rápido
- D Sí, pero no está malo
- Profesor ¿Cómo estará adentro?
- M Líquido
- Profesor Va estar líquido, ya veo

Las preguntas de los profesores

Las preguntas tienen un papel central en el discurso en el aula, tanto las preguntas formuladas por el profesor, como aquellas hechas por los estudiantes entre sí y al (la) profesor(a). El preguntar ocupa una gran parte de la conversación de los profesores y es uno de los factores más importantes en la determinación de las oportunidades de los estudiantes para el desarrollo de la comprensión a través de la indagación. No es la frecuencia de las preguntas lo que importa, sino su forma y contenido, y cómo aparecen en los patrones de discurso en el aula.

Respecto a la *forma*, las distinciones más importantes son entre las preguntas "abiertas" y las preguntas "cerradas" y entre las preguntas "centradas en la materia" o "centradas en la persona". Las preguntas abiertas permiten a los estudiantes expresar su visión u observación ("¿Qué notas acerca de...?") en lugar de responder a un punto concreto planteado por el profesor ("Son todos estos del mismo tamaño?"). Las preguntas centradas en la materia preguntan directamente sobre el tema ("¿Por qué éste... toma más tiempo que ...?"), a la vez que las preguntas centradas en la persona piden las ideas de los estudiantes ("¿Por qué crees que esto ... toma más tiempo que ...?"). Es más esperable que las preguntas abiertas y centradas en la persona entreguen más información a los profesores sobre lo que los estudiantes están observando y pensando y que podría ser importante para decidir cómo ayudarlos.

En relación con el *contenido*, las preguntas deben ser consistentes con el propósito. No debemos hacer preguntas sin una razón y sin interés en la respuesta. Para que la respuesta sea útil, debe entregar la información o estimular el tipo de respuesta requerida. Las preguntas se utilizan en todas las partes del ciclo de la evaluación formativa (Figura 1 página 20), pero, en el contexto de ECBI, muy especialmente en revelar y ayudar a los estudiantes a tomar los próximos pasos en el desarrollo de sus ideas y habilidades y en el fomento de la colaboración, el intercambio de ideas, la reflexión y evaluación. El contenido será diferente de acuerdo con el enfoque en un momento en particular. Por ejemplo preguntas tales como:

¿Cuál piensas es la razón para...?

¿Por qué crees que esto ocurre?

Preguntan directamente por las ideas de los estudiantes sobre cómo se pueden explicar las cosas, mientras que:

¿Qué observan que está pasando aquí?

¿Qué creen pasaría si....?

Pueden contestarse sin dar una explicación.

Las preguntas que se relacionan a las habilidades de indagación se formularán según las habilidades pertinentes en una situación particular. Una pregunta diseñada para estimular a los estudiantes a hacer una predicción podría ser "¿Qué crees que hará que esta planta crezca más rápido?", mientras que para fomentar la interpretación de los datos podría ser "¿Qué piensas hizo que esta planta creciera más rápido?" Este tipo de preguntas, que fomentan el uso de habilidades de investigación científica, deben ser distinguidas de aquellas que preguntan acerca de las ideas desarrolladas a partir de la indagación o las destinadas a fomentar la aplicación de las ideas a situaciones distintas de las investigadas.

Algunos ejemplos de preguntas para distintos propósitos, utilizando como contexto algunas actividades de los estudiantes que investigan las sombras, se presentan los Recuadros de 12, 13 y 14.

61 Quoted from Harlen, W. and Qualter, A. (2009) *The Teaching of Science in Primary Schools*. 5th edn London: Routledge, p 100-101

Una buena forma de comenzar para acceder a las ideas de los estudiantes es mediante el uso de preguntas abiertas. Si esto sólo conduce a una respuesta vaga, puede ser necesario seguir con preguntas más focalizadas, pero centradas en la persona. Para ayudar a los estudiantes a tomar los próximos pasos, las preguntas útiles deben seguir las ideas que se han expresado y estar diseñadas para alentar a los estudiantes a relacionar diferentes observaciones para ver si sus ideas funcionan en otras circunstancias. También puede estimular a los estudiantes a considerar ideas diferentes, generando un "andamiaje" de alternativas. Estas preguntas pueden ser bastante menos abiertas que las preguntas destinadas a poner en evidencia las ideas.

Recuadro 12: Preguntas para descubrir y estimular el desarrollo de ideas

Descubriendo las ideas

- ¿Qué piensas genera la sombra?
- ¿Por qué piensas que estos objetos dan sombras más oscuras que estos otros?
- ¿Cómo explicas la forma de una sombra?

Ayudando a los próximos pasos

- ¿Cómo explicaría tu idea el hecho que el color de la sombra sea igual para todos los objetos?
- ¿Qué otra idea podría explicar la forma de la sombra?
- Si el objeto elimina la luz de la pared, cómo piensas que esto explicaría que la sombra sea más grande si el objeto está más cerca de la linterna?

Las respuestas a preguntas que buscan averiguar acerca de las habilidades de indagación de los estudiantes, en el Recuadro 13, pueden ser tanto acciones como palabras, ofreciendo a la profesora la oportunidad de ver cuáles son las habilidades que los alumnos ya tienen. Las preguntas sugeridas para descubrir lo que los alumnos saben hacer también proporcionan oportunidades para el desarrollo de habilidades de indagación. Una vez más, están abiertas y redactadas para alentar a los estudiantes a responder sin buscar una respuesta "correcta". Algunas pueden involucrar andamiaje, proporcionando apoyo para pensar a lo largo de ciertas líneas, como por ejemplo acerca de las variables o relacionando lo que encontraron con su pregunta inicial.

Las preguntas en el Recuadro 14 se han diseñado para estimular a los estudiantes a trabajar en una cooperación genuina y no como individuos (aún cuando trabajen en grupos o en parejas) o en competencia unos con otros. Ellas requieren una respuesta que refleje el pensamiento combinado, que puede ser sobre cómo explicar algunas observaciones o cómo planificar una investigación, etc. Las preguntas para la reflexión y la evaluación tienen como propósito asegurar que los estudiantes reproduzcan en sus mentes lo que han hecho y tomen conciencia sobre cómo sus ideas han cambiado. Sin esta reflexión, sus ideas tienden a revertir a su anterior forma de pensar. Estas preguntas requieren que los niños hablen sobre lo que han aprendido y cómo lo han aprendido, es decir, que "aprendan a aprender" así como sobre las cosas que han investigado.

Recuadro 13: Preguntas para descubrir y estimular el desarrollo de habilidades de indagación

Descubriendo

- ¿Qué te gustaría descubrir sobre las sombras? (hacer preguntas)
- ¿Qué crees pasaría si movemos los objeto de esta forma? (predicción)
- ¿Qué podrías hacer para encontrar lo que afecta el tamaño de la sombra? (investigación)
- ¿Qué has encontrado sobre si hay una conexión entre la posición de la linterna y el tamaño de la sombra? (interpretación)

Ayudando a los próximos pasos

- ¿Qué encuentras si mides el tamaño de la sombra antes y después de mover la linterna?
- Si la sombra se hace más grande cuando se mueve la linterna de esta manera, ¿qué crees que pasará si la mueves en la otra dirección?
- ¿Cómo te asegurarás que es la posición de la linterna y no otra cosa lo que hace la diferencia?
- ¿Qué has descubierto acerca de cómo se puede cambiar el tamaño de la sombra?

Recuadro 14: Preguntas para estimular la colaboración, compartir ideas, la reflexión y la evaluación

Colaboración y compartir ideas

- ¿Cuántas ideas diferentes puede sugerir su grupo para explicar lo que encontraron?
- Después de usar diferentes ideas para explicar lo que encontraste, ¿hay una que parece mejor?
- De todas las ideas, ¿cuáles podrían someterse a prueba?
- ¿Cuál es el acuerdo a que llegaron respecto a la forma de encontrar qué idea funciona?
- ¿Qué haría cada persona en el grupo en esta investigación?

Reflexión y evaluación

- ¿Qué encontraron que no sabían antes?
- ¿Han cambiado su idea sobre...?
- ¿Qué les hizo cambiar de idea?

¿Qué es lo que no entienden de ... ?

- ¿Hay algo que todavía quieran encontrar?
- Si lo hicieran otra vez, ¿qué cambiarían de manera que pudieran aprender más?

Tiempo para responder

Las preguntas bien formuladas merecen respuestas bien pensadas y los estudiantes necesitan disponer de tiempo para responder a la pregunta. Las preguntas sugeridas en los Recuadros 12, 13 y 14 están diseñados para provocar el pensamiento; ellas requieren una respuesta bien pensada. Existe, posiblemente, un lugar para la prueba de memoria de respuesta muy rápida, o para los quiz, pero eso no es lo que nos interesa en el contexto del aprendizaje a través de la indagación. La presión para responder rápidamente reduce el valor de preguntar para los propósitos que hemos estado discutiendo. Por ello es necesario indicar a los niños que se requiere una respuesta pensada, no una respuesta rápida. Esto se puede hacer de varias maneras.

- La primera es aumentar el "tiempo de espera", el tiempo entre hacer una pregunta y recibir una respuesta. Los profesores suelen esperar una respuesta muy rápida y con ello disuaden a los estudiantes de pensar. La conocida investigación de Budd-Rowe⁶² mostró que extender el tiempo que los profesores esperan a los estudiantes que contesten tiene un marcado efecto en la calidad de las respuestas. Ella encontró que los maestros esperaban en promedio menos de un segundo después de hacer una pregunta, y si no había respuesta, rephraseaban o daban una pista o hacían una pregunta más fácil. Cuando se les pidió a los profesores aumentar el tiempo de espera a ocho o nueve segundos, la calidad de las respuestas de los estudiantes aumentó dramáticamente.
- La segunda es evitar reformular una pregunta, si ésta no es respondida rápidamente. Reformular la pregunta inevitablemente hace que sea más cerrada y menos útil. Cuando esto sucede regularmente los estudiantes se dan cuenta de que, si esperan, el profesor hará una pregunta más sencilla, a menudo indicando la respuesta esperada.

Algunos profesores consideran que es mejor no permitir que los estudiantes levanten la mano para responder a este tipo de preguntas reflexivas. Esperan que todo el mundo sea capaz de responder con suficiente tiempo para pensar. De esta manera el profesor da tiempo para pensar y luego llama a los estudiantes por su nombre para contribuir. Esto indica que lo que se valora es el pensamiento, no la velocidad, y estimula a todos los estudiantes a prestar atención a la pregunta y así involucrarse más en escuchar las respuestas de otros estudiantes.

- Otra estrategia adecuada para algunas preguntas y situaciones, es sugerir que los estudiantes discutan sus respuestas con un compañero o un grupo por dos o tres minutos antes que el profesor solicite las contribuciones.

Retroalimentación a los estudiantes

La retroalimentación ha sido descrita como "una de las más poderosas influencias sobre el aprendizaje y el logro", pero con la advertencia agregada de que "este impacto puede ser positivo o negativo".⁶³ Tiene un papel clave en la evaluación formativa, ya que es el mecanismo por el cual las futuras oportunidades de aprendizaje se ven afectadas por el aprendizaje previo. La retroalimentación es más evidentemente entregada por los profesores a los estudiantes por vía oral o por escrito, pero también, quizás inconscientemente, por el gesto, la entonación y desde luego, por la acción, como en la asignación de tareas a los estudiantes. El cómo entregan retroalimentación los profesores y hacia qué se enfoca, está influenciado por su visión del aprendizaje (véase el Capítulo 4). La visión constructivista del aprendizaje conduce a una interacción entre el (la) profesor(a) y los estudiantes en la cual los estudiantes responden a los comentarios y sugerencias de los profesores en

62 Budd-Rowe, M. (1974) Relation of wait-time and rewards to the development of language, logic and fate control: Part II, *Journal of Research in Science Teaching*, 11(4) 291-308

63 Hattie, J. and Timperley, H. (2007) The power of feedback. *Review of Educational Research*, 77, 81-112.

lugar de la comunicación de un solo lado - desde el(la) profesor(a) al estudiante - que es típica de una visión conductista del aprendizaje.

Hay dos aspectos principales de la retroalimentación a los estudiantes a considerar - la forma y el contenido.

La forma de la retroalimentación

Ha sido tradicional que la retroalimentación sobre el trabajo de los estudiantes se entregue en la forma de una nota, u otro signo del juicio de su idoneidad. Un estudio de Butler⁶⁴, que es frecuentemente citado, compara la retroalimentación entregada en la forma de una nota con dos otras formas, una que entrega comentarios sobre el trabajo y sobre cómo mejorarlo, y otra que entrega comentarios y notas. Se trató de un estudio bien diseñado y complejo que involucró distintos tipos de tareas y estudiantes con diferentes niveles de logro. Uno de los resultados obtenidos fue que entregar retroalimentación solo con comentarios llevó a mayores logros para todos los estudiantes y para todas las tareas. Un resultado interesante fue que entregar retroalimentación con nota y comentarios no era más efectivo que notas por si solas. Parece ser que los estudiantes se apropian de las notas y hacen caso omiso de los comentarios que las acompañan. Ellos recurren a las notas buscando un juicio más que una ayuda para avanzar en el aprendizaje. Cuando no se entrega una nota ellos se involucran con lo que el profesor desea traer a su atención. Los comentarios, entonces, tienen la oportunidad de mejorar el aprendizaje de la forma prevista por el (la) profesor(a). Para que se cumpla este propósito, los comentarios deben ser positivos, sin juicio y en lo posible identificar los siguientes pasos.

En su trabajo sobre la implementación de la evaluación formativa, Black et al. (2003) introdujo a los profesores a la investigación de Butler y a los resultados de la observación en sus propias aulas, lo que mostró que:

- Los estudiantes rara vez leyeron los comentarios, prefiriendo comparar sus notas con sus compañeros como la primera reacción al recibir su trabajo de vuelta
- Los profesores rara vez le dan a los estudiantes tiempo en la clase para leer los comentarios que se escriben sobre su trabajo y, probablemente pocos, si alguno, de los estudiantes los vuelven a considerar en su casa
- A menudo los comentarios son breves y/o no específicos, por ejemplo, "¿Detalles?"
- Las mismas observaciones escritas aparecen con frecuencia en el cuaderno del alumno, implicando que los estudiantes no parecen tomar nota de ellas o actuar considerando los comentarios.⁶⁵

Cuando se les preguntó a los estudiantes sobre la forma en que se evaluaban sus cuadernos, dijeron que no querían que los profesores usaran lápiz rojo, y que esperaban que escribieran en forma legible y que hicieran comentarios comprensibles.

El contenido de la retroalimentación

El punto principal que emerge tanto de los estudios de investigación, como de la experiencia de la práctica efectiva es una distinción entre una retroalimentación que entrega información y una retroalimentación que entrega un juicio. Esto aplica a los comentarios orales tanto como a los escritos.

64 Butler, R. (1988) Enhancing and undermining intrinsic motivation: the effects of task-involving and ego-involving evaluation on interest and performance, *British Journal of Educational Psychology* 58, 1-14

65 Black et al, (2003) op cit p 43

Comentarios que dan información:

- se focalizan en la tarea y no en la persona
- alientan a los estudiantes a pensar sobre el trabajo no sobre cuán "buenos" son ellos
- indica qué hacer a continuación y da ideas sobre cómo hacerlo.

Retroalimentación que entrega un juicio:

- se expresa en términos de cuán bien lo ha hecho *el alumno* (lo que incluye alabanza y crítica) en lugar de lo bien que se ha hecho el *trabajo*
- entrega un juicio que estimula a los estudiantes a auto-etiquetarse
- proporciona una nota que los estudiantes utilizan para compararse con los demás

Un aspecto adicional de la retroalimentación efectiva es que focaliza la atención del estudiante en lo que es relevante para el logro de determinados objetivos de la lección. Así, si el objetivo principal de una indagación es desarrollar habilidades de indagación científica, entonces la retroalimentación debe ser sobre ese aspecto en lugar de, por ejemplo, los resultados concretos obtenidos.

Al inicio de su trabajo con los profesores, Black et al. notaron que los comentarios formulados por los profesores sobre los estudiantes eran ya fuese una evaluación general ('Bueno', 'Bien hecho') o se centraban en la presentación o en la completitud del trabajo. Uno de los profesores, reconociendo que estos comentarios no ayudan el aprendizaje de los estudiantes", escribió en sus notas:

Un comentario blando y poco útil como "Buen trabajo, Jaspaul, esto está mucho más ordenado y parece demostrar que has hecho un importante esfuerzo" no mostrará un cambio significativo en el logro, porque no dice nada sobre el aprendizaje del individuo. No hay un objetivo y el estudiante, aunque esté consciente de que el profesor está contento con él o ella, no podrá ser culpado por pensar que la pulcritud lo es todo ... Los estudiantes no se dan fácilmente cuenta de lo que están aprendiendo, a menudo porque nosotros, como profesores, no se lo decimos de una manera apropiada.⁶⁶

Comentarios que ayudan al aprendizaje son:

"Usted ha identificado correctamente las sustancias químicas que son elementos y cuáles son compuestos, ahora traten de dar una explicación general sobre la diferencia entre los elementos y los compuestos."

"Miren hacia atrás en sus notas sobre la estructura de la célula y luego vean si están de acuerdo con lo que han escrito aquí".

"¿Qué has descubierto que te hizo decidir que la masa de los objetos no hace ninguna diferencia en la rapidez con que caen?"

Conclusiones sobre la retroalimentación a los estudiantes

Recomendaciones que vienen de la investigación dicen que si se espera que la retroalimentación ayude el aprendizaje, ésta:

- Debería ser en la forma de comentarios sin notas o puntajes

66 Black et al, (2003) op cit p 45

- Ya sea en forma oral o escrita, los comentarios sobre el trabajo de los estudiantes deben identificar lo que se ha hecho bien, lo que podría mejorarse y cómo podría mejorarse
- Los comentarios deben ayudar a los estudiantes a tomar conciencia de lo que han aprendido.
- Los profesores deben verificar que los alumnos comprenden sus comentarios.
- El tiempo debe ser planificado para que los estudiantes lean y, si es apropiado, respondan a los comentarios.

Retroalimentación hacia la enseñanza

La evaluación formativa se trata tanto de entregar información a los profesores como de la retroalimentación a los estudiantes. Las dos están estrechamente relacionadas, ya que el cómo los alumnos responden a las preguntas y a la retroalimentación de sus profesores y de otros estudiantes es una fuente de evidencia que el profesor puede utilizar para tomar decisiones sobre los próximos pasos de los estudiantes. Dado que todos los alumnos son individuos, no hay un único camino a seguir para asegurar el aprendizaje. Los profesores tienen que juzgar el valor de una intervención a partir del impacto de sus preguntas y otras acciones. En la recolección de datos relevantes para informar sus intervenciones, los profesores tienen que tener muy claros los objetivos que quieren que alcancen sus alumnos. De esta manera podrán distinguir los datos significativos de la amplia gama que potencialmente podría ser utilizada. Con este foco, ellos podrán recoger datos cuando los estudiantes participen en las investigaciones, observando, preguntando, escuchando cómo los estudiantes están utilizando las palabras y estudiando sus cuadernos. Una fuente importante de retroalimentación para la profesora viene de la autoevaluación de los alumnos y de la evaluación por pares que se discute más adelante, ya que si los estudiantes no tienen claro lo que deben lograr a través de su trabajo, se pueden estar utilizando criterios inadecuados para juzgar el éxito.

Esta retroalimentación informa las decisiones acerca de si intervenir o cómo intervenir en el curso de las actividades de los alumnos. El proceso es cíclico y cada decisión cambia la situación. No todas las intervenciones tendrán el efecto positivo deseado; lo que sucede en las aulas no siempre – o quizás raramente – ocurre según lo previsto. La retroalimentación que los profesores reciben de las reacciones de sus estudiantes les permite probar algo diferente, si es necesario, con el fin de ayudar a los estudiantes a progresar. Puede ser necesario que un profesor cambie los planes cuando los estudiantes están teniendo dificultades en lugar de arriesgarse a una sensación de fracaso. De esta manera la retroalimentación permite a los profesores regular la enseñanza para maximizar el aprendizaje.

Autoevaluación y evaluación entre pares de los estudiantes

Un objetivo común de la evaluación formativa y ECBI es que los estudiantes se vuelvan cada vez más capaces de tomar parte en las decisiones sobre la calidad de su trabajo y desarrollen su comprensión sobre qué está involucrado en el aprendizaje. Los estudiantes son, en cualquier caso, responsables de su aprendizaje, pero *el que tomen responsabilidad* por éste dependerá de su participación en las decisiones. Esta participación está representada por las flechas de dos cabezas en la Figura 1 en la página 20.

Los estudiantes, al igual que todos los alumnos, dirigen su esfuerzo con mayor eficacia si saben lo que están tratando de lograr, más que sólo sabiendo lo que tienen que hacer. Los profesores son buenos para decirles a los estudiantes lo que tienen que hacer ("envolver los cubitos de hielo en diferentes materiales y observar cuál toma más tiempo en fundirse"), pero no tan buenos en otorgar un propósito y un objetivo ("para ver si algunos materiales son mejores que otros en evitar que el hielo se derrita y tratar de explicar lo que encuentras"). Un requisito previo para poder juzgar su

trabajo es que los estudiantes entiendan lo que están tratando de hacer, no en términos de lo que se ha encontrado, sino en términos de la pregunta que debe contestarse o el problema a resolver.

Por supuesto que cuando la pregunta o el problema han sido planteados por los estudiantes, no se hace necesario comunicarlo. Pero esto será siempre sólo una de las diferentes prácticas pedagógicas utilizadas por docentes, que incluyen actividades dirigidas por el(la) profesor(a). Siempre habrá situaciones en las que los profesores introducen la pregunta y, aunque usen habilidades para ayudar a los estudiantes a que "las hagan suyas", los profesores tendrán que asegurarse de que los estudiantes comprendan el propósito y el objetivo de la actividad. El indicar los objetivos en el inicio de una lección no es la única forma de transmitirlos, ni necesariamente la mejor. La comprensión de los objetivos, o el por qué están trabajando de una manera particular, se puede reforzarse a través del diálogo y de preguntas durante la actividad y en la discusión de lo que se hizo y se encontró.

Con el fin de evaluar su trabajo, los estudiantes no sólo necesitan conocer el propósito de lo que están haciendo, sino que necesitan tener alguna noción del estándar al cual deben aspirar, esto es, qué es "un buen trabajo" en un contexto particular. Algo de esto se transmite implícitamente a través de la retroalimentación que los profesores dan a los estudiantes. Los profesores también pueden discutir en forma más explícita qué hace que una pieza de trabajo o investigación sea mejor que otra, utilizando ejemplos recogidos con esa finalidad y de forma anónima. Por otra parte, los ejemplos de las colecciones publicadas o creadas en la escuela para ayudar a los profesores a evaluar el trabajo, podrían ser compartidos con los estudiantes (ver más adelante). En el Recuadro 15 se describen algunos ejemplos de cómo los profesores de estudiantes de primaria se han acercado a la discusión de lo que es un buen trabajo⁶⁷.

67 Adapted from Harlen, W. (2006) Teaching, Learning and Assessing Science 5-12. London: Sage. p171

*Recuadro15: Comunicando a estudiantes de enseñanza básica los criterios para evaluar su trabajo*⁶⁷

Usando ejemplos

Una profesora de alumnos de 10 años, gastó algún tiempo al inicio del año escolar para discutir con su clase qué se espera de un "buen" informe de una investigación científica. Ella les dio a los estudiantes dos ejemplos anónimos de textos escritos por estudiantes sobre investigaciones realizadas en años anteriores. Uno era un relato claro y bien expuesto de modo que el lector pudiese comprender lo que se había hecho, aunque la escritura era irregular y había algunas palabras con errores ortográficos. Había diagramas para ayudar al relato, con rótulos. Los resultados estaban en una tabla, y el autor había expresado lo que él o ella pensaba que querían decir, admitiendo que los resultados no respondían completamente a la pregunta inicial. Había un comentario sobre cómo se podrían haber mejorado las cosas. El otro relato era ordenado, atractivo a la vista (los diagramas estaban coloreados, pero no rotulados), pero en el contenido no mostraba ninguna de las características que se observaban en el otro texto.

La profesora pidió a los estudiantes que trabajaran en grupos para comparar sus trabajos y que listaran las características buenas y malas de cada uno. Luego se les pidió que dijeran cuáles pensaban eran las cosas más importantes que debía contener un "buen" informe. La profesora reunió todas las ideas y añadió algunos puntos por su cuenta, a lo que los estudiantes habían acordado. Más tarde hizo copias para que todos los estudiantes las guardaran en sus carpetas de ciencia. Pero ella también continuó explorando con ellos cómo llevar adelante una investigación con el fin de poder escribir un buen informe. Estos puntos también fueron reunidos utilizando las palabras de los niños y se imprimieron para ellos.

Lluvia de ideas

Una variante de lo anterior es una lluvia de ideas, por ejemplo, acerca de cómo llevar a cabo una investigación específica de modo que los niños estén seguros del resultado. La lista de aquello que debe tenerse en cuenta, se puede convertir preguntas (¿Mantuvimos todo igual excepto por...? ¿Cambiamos ...? ¿Observamos ...? ¿Comprobamos los resultados? etc.) Antes de terminar su investigación ellos chequean su lista, la que se convierte en una herramienta de autoevaluación para esa pieza de trabajo.

Discutiendo el "mejor trabajo"

Este enfoque se puede utilizar con los estudiantes desde la edad de 8 años. Se inicia con los estudiantes seleccionando "su mejor trabajo" el que colocan en una carpeta. Se deja un tiempo para que el profesor converse con cada estudiante sobre por qué han seleccionado estos trabajos. Durante esta discusión se pone en evidencia la forma en que los estudiantes están juzgando la calidad de su trabajo. Estas se aceptan sin comentario, independientemente de si reflejan la visión del profesor sobre un buen trabajo o no. Para aclarar los criterios que utilizan los estudiantes, el profesor podría preguntar. "Dime qué es lo que particularmente te gustó de este trabajo?" Gradualmente se podrá sugerir criterios sin forzar lo que los estudiantes deberían seleccionar. Esto se puede realizar a través de los comentarios que se hagan sobre su trabajo. "Esta fue una muy buena forma de mostrar tus resultados, pude ver fácilmente cuál era el mejor". "Me alegra que pienses que esta fue tu mejor investigación, porque si bien no obtuviste el resultado esperado, lo hiciste con mucho cuidado y te aseguraste que el resultado fuese válido".

Con estudiantes mayores se pueden usar enfoques más directos y sofisticados para fomentar la auto-evaluación. El Recuadro 16 describe cómo un profesor de biología ayudó a sus estudiantes de 16 años a comprender la forma de evaluar su trabajo. Esto surgió en el curso de ayudar a los estudiantes a aclarar sus ideas sobre la nutrición de las plantas.

*Recuadro 16: Estudiantes secundarios utilizando autoevaluación para mejorar su trabajo*⁶⁸

Los estudiantes recibieron la siguiente pregunta para discutir: "Si un villano se apodera de un producto químico que pudiese destruir la clorofila, ¿qué efecto tendría esto en las plantas en caso de ser vertido?" Se le pidió a cada grupo, formado por cuatro estudiantes, que escribieran entre tres y cinco criterios que pensaban debían cumplirse para el caso de una buena respuesta escrita. Estos criterios fueron discutidos por toda la clase y se elaboró la lista final :

- Diga lo que hace la fotosíntesis
- Explique la fotosíntesis como un proceso
- Coloque la ecuación para la fotosíntesis
- Describa el efecto de la ausencia de clorofila en las plantas
- Agregue los efectos secundarios que tendría detener la fotosíntesis

Luego, los estudiantes escribieron sus respuestas como tarea en su casa, las que el profesor evaluó escribiendo comentarios sobre su trabajo. En seguida, en parejas, los alumnos leyeron los comentarios y el trabajo de cada uno, para comprobar que entendían lo que la profesora les pedía mejorar. A continuación se les dio el tiempo en la lección para volver a redactar y mejorar sus respuestas.

Evaluación por pares

La importancia de la evaluación por pares para la evaluación formativa ha sido reconocida, tras su fuerte apoyo por Sadler.⁶⁹ Black ofrece varios argumentos en favor de alentar a los estudiantes a juzgar el trabajo de los demás. Algunos de estos se basan en ayudar a los estudiantes a comprender mejor los objetivos y criterios de calidad al ver el trabajo de otros en lugar de los suyos.⁷⁰ Algunas de las razones se basan en consideraciones prácticas – como que los estudiantes se motivarán a realizar un mejor trabajo si saben que éste va a ser leído por otro estudiante. Sin embargo, como otros investigadores han señalado, estos argumentos no tienen en cuenta las relaciones de poder asociadas con la situación de una persona juzgando el trabajo de otro.⁷¹ Los estudios de caso de Crossouard sobre la evaluación por pares de alumnos de 11 y 12 años, proporciona evidencia preocupante sobre cómo el género, la clase social y las jerarquías de logro de los estudiantes están "implicadas en procesos que suponen "neutralidad" de los juicios evaluativos."⁷² Se ha observado que los beneficios de la evaluación por pares no tiene el mismo efecto para todos, ya que puede

68 Based on Black et al (2003) op cit p 63

69 Sadler (1989) op cit

70 Black et al (2003) op cit p 50

71 Crossouard (2012) Absent presences: the recognition of social class and gender dimensions within peer assessment interactions, *British Educational Research Journal*, 38 (5) 731-748

Pryor, J. and Lubisi, C. (2001) Reconceptualising educational assessment in South Africa –testing times for teachers, *International Journal for Educational Development*, 22 (6), 673-686

72 Crossouard (2012) op cit p 736

apoyar a ciertos estudiantes, mientras puede resultar “opresiva” para otros. Así, los profesores podrían necesitar ayuda para reconocer los problemas de equidad que se plantean y para hacer frente a la influencia de la clase social, el género y la habilidad general cuando se practica la evaluación por pares

Capítulo 6

Implementación de la evaluación sumativa de ECBI

Después de considerar, en el Capítulo 5, algunos enfoques para la puesta en práctica de la evaluación formativa para apoyar el aprendizaje de los estudiantes a través de la indagación, en este capítulo veremos algunos métodos de aplicación de la evaluación sumativa. El objetivo es proporcionar información sobre lo que los estudiantes comprenden y pueden hacer en un momento determinado, con énfasis en los conocimientos científicos y las habilidades de indagación científica. En contraste con los requisitos de la evaluación formativa, la confiabilidad es importante en la evaluación sumativa, ya que los resultados pueden ser utilizados para comparar o seleccionar a los estudiantes y también pueden ser utilizados como indicadores de la eficacia docente y la escuela. Sin embargo, la validez también es importante y, dada la interacción entre la validez y la confiabilidad señalada en el Capítulo 1 (Recuadro 2), un factor clave en la elección de los métodos de evaluación es asegurar que no se comprometa la validez en el esfuerzo por alcanzar una exactitud aparente. Aquí nos centramos en la evaluación sumativa de la comprensión científica y las habilidades de indagación científica, ya que éstos constituyen el mayor desafío y su descuido en la evaluación es la amenaza más grande para ECBI.

Centrarse en la evaluación sumativa no significa que otros logros, tales como el conocimiento de hechos, el vocabulario científico y el uso de las convenciones no han de ser evaluadas, pero ya existen muchas formas de hacerlo - por ejemplo, a través de tests breves desarrollados por el profesor en momentos apropiados. Más desafiante es la evaluación del progreso de la comprensión y las competencias que se intenta desarrollar a través de ECBI. Si estas no se incluyen en la evaluación, existe el peligro real de que sean descuidadas en la enseñanza.

Evaluación de la comprensión y de las habilidades científicas

Comprensión y “grandes” ideas

Para empezar debemos tener en cuenta qué es la comprensión y cómo se relaciona con los conocimientos específicos y generales. La comprensión no es fácil de precisar, como señala White:

(comprensión) es una función continua del conocimiento de una persona, no es una dicotomía y no es necesariamente lineal. Decir si alguien comprende es un juicio subjetivo que varía con el juez y con el status de la persona que está siendo juzgada. El conocimiento varía en cuanto a su relevancia para la comprensión, pero esta relevancia es también un juicio subjetivo⁷³

La comprensión existe en varios niveles y en las diferentes formas de explicar los fenómenos. La comprensión de un estudiante de escuela básica y de un estudiante de ciencias de educación superior será diferente, pero ambos pueden cumplir con las expectativas propias de su etapa de aprendizaje. Por ejemplo, un joven estudiante podría explicar el fenómeno de la disolución en términos de lo que sucede cuando algunos sólidos se adicionan a algunos líquidos. Se espera que los estudiantes mayores lo expliquen en términos del comportamiento molecular que se aplica a las soluciones de gases y de

73 White, R.T. (1988) *Learning Science*. Oxford: Blackwell. p 52

líquidos en líquidos. Lo que probablemente sea lo mismo para ambos, es que las ideas que tienen les hacen sentido y se corresponden con las experiencias que están tratando de entender en un momento determinado. A medida que las experiencias de los estudiantes más jóvenes se expanden y sus ideas ya no proporcionan una explicación satisfactoria, entonces el fenómeno necesitará ser entendido de una manera diferente. La comprensión progresiva acompañada de experiencias amplificadoras también puede ser vista como el desarrollo sucesivo de ideas "más grandes", vinculando más fenómenos y siendo más poderosas para explicar las cosas.

Las "grandes" ideas son las que se pueden aplicar en diferentes contextos, ellas permiten a los estudiantes dar sentido a una amplia gama de fenómenos mediante la identificación de vínculos esenciales o "patrones significativos" entre los diferentes eventos, objetos o fenómenos sin ser desviados por las diferencias en las características superficiales. Pueden ser ideas científicas (tales como fuerza y movimiento) o sobre los procedimientos científicos (como la manipulación de variables). Simplemente memorizar hechos o un conjunto de procedimientos no es compatible con esta capacidad de aplicar lo aprendido a contextos más allá de aquellos en que se han aprendido. El conocimiento que es comprendido, es el conocimiento útil que se puede utilizar en la resolución de problemas y en la toma de decisiones.

Así, el progreso en la comprensión no se evidencia en la cantidad de conocimiento, sino en si se organiza en la mente, dando forma a ideas "grandes", de manera que se pueda acceder a él fácilmente y ser aplicado a nuevas experiencias. Esta demostración de la comprensión es un aspecto clave de desempeño que distingue a los "expertos" de los "novatos".

Los expertos han podido almacenar gran cantidad de conocimientos y habilidades en un dominio particular. Pero quizás lo más importante, han organizado este conocimiento en sus mentes de manera que lo hacen más recuperable y útil.

... Estos métodos de codificación y organización ayudan a los expertos a interpretar la nueva información y a reconocer características y patrones significativos de la información que pueden ser pasados por alto por aprendices menos competentes. Estos esquemas también permiten a los expertos recuperar los aspectos relevantes de su conocimiento cuando se enfrentan a un problema.⁷⁴

Aquí, entonces, se entregan algunas claves sobre la evaluación en ciencias que permiten realizar inferencias acerca de la comprensión de los estudiantes:

- La tarea de evaluación tiene que incluir la aplicación del conocimiento, no sólo recuerdo.
- Esto significa que la tarea tiene que ser nueva, pero al mismo tiempo no debe estar tan lejos de su experiencia que no tenga ningún significado para los estudiantes.
- Reconociendo que la ciencia es sobre el mundo real, no uno ficticio, la tarea debe ser auténtica, sobre cosas y datos reales.

Hay otros requerimientos también, en el caso que la evaluación tenga que ser presentada dentro de un plazo razonable y no depender demasiado de competencias (como las habilidades de lectura y la escritura) distintas de la comprensión que se evalúa.

Habilidades de indagación científica

En todas las tareas de evaluación contextualizadas y relacionadas con temas que son importantes para los estudiantes, el desempeño se ve afectado por la familiaridad que los alumnos tienen con el contexto y la materia. Todas las competencias son utilizadas en un contexto y en relación con un cierto

74 Pellegrino et al op cit p 72/73

contenido. Aunque el contexto también influye en la aplicación de conceptos científicos, presenta problemas particulares cuando el foco de la pregunta se centra en obtener datos sobre el uso de habilidades de indagación científica.

Las preguntas y las tareas tienen que ser hechas en función de *algo*, las observaciones se hacen acerca de objetos y eventos particulares, las investigaciones son planificadas para responder a preguntas sobre fenómenos particulares, tiene que haber algún contenido involucrado cuando se utilizan las habilidades. El contenido hará una diferencia respecto a la utilización de las habilidades. Un estudiante puede ser capaz de planificar adecuadamente una investigación acerca de una situación donde él/ella tiene algún grado de conocimiento de cuáles son las variables a controlar, pero fracasar en hacerlo si el tema es desconocido. Esto tiene importantes consecuencias para la evaluación. El tema de una tarea o ítem de una prueba en particular es sólo una de las alternativas de contenidos potenciales. En teoría, el resultado de un estudiante sería diferente en el caso de haber elegido una materia diferente, así que hay una variación, o error, en los resultados relacionados con la elección. Esto se describe como el "error de muestreo". Es un error inevitable, ya que no hay dos tareas con contenidos o contextos distintos que puedan ser exactamente equivalentes.

De esto se deducen algunos puntos sobre la evaluación de habilidades de investigación científica:

- En primer lugar, y muy obvio, que los estudiantes necesitan estar involucrados en el uso de las habilidades de indagación a fin de determinar lo que pueden hacer.
- En segundo lugar, dado que el contexto y el tema de la situación en la que los conocimientos se van a utilizar afecta la capacidad de utilizar las habilidades, la tarea debe establecerse, en lo posible, en un contexto familiar, o varios contextos deberían utilizarse para reducir el error de muestreo.
- En tercer lugar, como en el caso de la evaluación de la comprensión, las tareas deben ser auténticas y atractivas para los estudiantes.

La tarea puede involucrar

- realizar en la práctica una indagación completa para abordar una pregunta o problema dado, proporcionando oportunidad para obtener evidencia sobre el uso de una serie de habilidades
- elaborar un diseño, en papel, de una investigación completa para enfrentar una determinada pregunta o problema
- considerar una parte particular de una determinada investigación, tales como las variables que necesitan ser manipuladas o controladas, la evidencia que necesita ser recopilada, o la interpretación de algunos datos dados.

Métodos de evaluación sumativa de los objetivos de ECBI

Esta breve discusión de lo que se necesita para evaluar la comprensión y las habilidades de indagación científica muestra que en ambos casos los estudiantes deben estar trabajando en tareas en las que algunos aspectos de la indagación estén involucrados. Ambos también requieren alguna novedad en la tarea que están llevando a cabo los estudiantes de modo que estén usando sus conocimientos o habilidades, y no simplemente recordando información, razones o procedimientos que se han establecido en la memoria. Por ejemplo, la tarea de averiguar cómo el grosor de un alambre conductor afecta la resistencia no servirá para evaluar la capacidad de planificar y llevar a cabo una investigación, si es algo que los estudiantes ya han hecho más de una vez. Tampoco una pregunta acerca de por qué la edad de un árbol se puede estimar a partir de los anillos de crecimiento

servirá para estimar la comprensión de ideas sobre el crecimiento de los árboles, si la razón ha sido memorizada sin estar vinculada a estas ideas.

Para evaluar la comprensión, la tarea debería solicitar la explicación de un evento o la interpretación de datos o una predicción que implique la aplicación de algunos conceptos. Para evaluar habilidades científicas, la tarea debe involucrar el uso de una o varias habilidades indagatorias, tales como predecir, planificar, conducir una investigación o interpretar datos entregados. Sin embargo, como se indicó en la página 13, no es posible evaluar habilidades sin involucrar algún conocimiento propio de la asignatura. Al mismo tiempo, las tareas utilizadas para evaluar comprensión requerirán algún uso de habilidades (explicar, interpretar, predecir). Es así como siempre habrá algún aspecto de la comprensión y de las habilidades requerido en *todas* las tareas. Lo que determina si una tarea está esencialmente evaluando comprensión o una habilidad, será el nivel de demanda de uno u otro, y el crédito que se da a diferentes tipos de respuestas en el proceso de evaluación.

En cuanto a los métodos de evaluación de estos objetivos, consideramos en primer lugar lo que se puede hacer a través de las pruebas, ya que es el enfoque más común para la evaluación sumativa. Más tarde vemos alternativas a las pruebas.

El potencial y las limitaciones de las pruebas para evaluar los objetivos de ECBI

La evaluación con fines sumativos debe ser lo más confiable posible (libre de error y sesgo). Hay una tendencia a usar tareas especiales o tests, ya que pueden ser controlados y presentados a todos los estudiantes de la misma forma, por lo que parecen ofrecer a los estudiantes las mismas oportunidades de demostrar lo que pueden hacer. La realización de pruebas, como se señaló en el Capítulo 1, es un método de evaluación en el que los procedimientos, como la tarea a realizar y, a menudo las condiciones y tiempos para responder están especificadas. Las pruebas se califican (asigna una nota) normalmente usando un esquema prescrito (rúbrica), ya sea por el(la) profesor(a) de los estudiantes o examinadores externos, los que suelen ser profesores de otras escuelas, o por máquinas. La razón para los procedimientos uniformes es permitir la comparabilidad entre los resultados de los estudiantes, quienes pueden dar las pruebas en diferentes lugares. Hay diferentes formas de prueba de acuerdo con la naturaleza de la tarea y la forma en que se da una respuesta. De este modo hay

- pruebas de desempeño (a veces llamadas "prácticas")
- pruebas integradas (situadas en el contexto del trabajo regular)
- pruebas de opción múltiple (donde se ofrecen alternativas de respuesta)
- pruebas de respuesta abierta, (donde el estudiante escribe respuestas cortas o largas en sus propias palabras)
- pruebas con libro abierto (donde los estudiantes tienen acceso a un número controlado de fuentes)

... y muchas más. La mayoría implican algún escrito, excepto en algunas pruebas de desempeño y pruebas de lectura para los niños pequeños. Las pruebas más formales, que llevan a un certificado o cualificación, se describen a menudo como exámenes.

El potencial de los tests para evaluar conocimientos científicos y habilidades de investigación científicas se puede apreciar mejor a través de algunos ejemplos de pruebas escritas y pruebas de desempeño.

Preguntas escritas

Las características de las preguntas orales de buena calidad señaladas en el Capítulo 5 se aplican igualmente a las preguntas escritas. Esto es, con el fin de obtener acceso al pensamiento de los

estudiantes, es útil hacer preguntas en términos de lo que ellos piensan. Las preguntas "abiertas", pidiendo a los estudiantes responder con sus propias palabras también tienen más probabilidades de entregar información sobre lo que saben y entienden, a diferencia de las preguntas donde los alumnos eligen entre las respuestas dadas. Estas últimas están abiertas a la adivinación y dan pistas a los estudiantes a través de las palabras usadas. Sin embargo, las preguntas de respuesta abierta también tienen desventajas, especialmente para los estudiantes más jóvenes que tienen más dificultades para ser específicos escribiendo que hablando. Interpretar las respuestas escritas de los estudiantes también puede ser un problema, aunque ahora hay programas informáticos que pueden hacer esto con una fiabilidad considerable.⁷⁵ Además, es necesario dejar en claro a los estudiantes qué tipo de respuesta se requiere y con qué extensión. Por ejemplo, preguntar: "¿Cuáles son las principales diferencias entre X e Y?", puede dar lugar a una larga lista de diferencias, algunas relevantes, otras no y algunas inexactas. Es mejor dejar en claro qué tipo de respuesta se requiere, como en: "Escribe las que crees que son las tres diferencias más importantes entre X e Y que les permite sobrevivir en hábitats diferentes". A continuación, proveer espacios que indican límites a la extensión de las respuestas. Las ventajas y desventajas de diferentes formatos de preguntas hacen que se incluya generalmente una mezcla de ellas, el equilibrio entre cada tipo variando de acuerdo con los objetivos de la prueba, el foco particular de una pregunta y la edad de los estudiantes. Algunos puntos adicionales surgen de considerar algunos ejemplos.

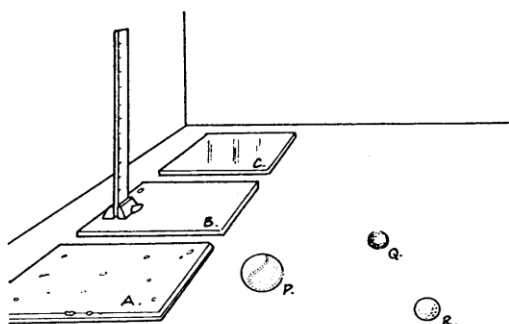
75 Streeeter, L. et al (2011) *Pearson's Automated Scoring of Writing, Speaking, and Mathematics*. Pearson

Ejemplo 1: Rebote de pelotas (Del Informe de Desempeño APU a la edad de 11 años)⁷⁶

Emma y Anita andaban buscando si la superficie en la cual una pelota rebota hace alguna diferencia en la altura del rebote.

Ellas encontraron tres diferentes superficies, que denominaron A, B y C.

Ellas también tenían tres pelotas diferentes P, Q y R.



Para una prueba imparcial, ¿qué deberían cambiar en sus ensayos, y qué deberían mantener igual?

Marque con una cruz Cambio o No Cambio para cada una de las siguientes cosas:

	<u>Cambio</u>	<u>No cambio</u>
La pelota	☐	☐
La superficie	☐	☐
La altura de lanzamiento	☐	☐

Comentario

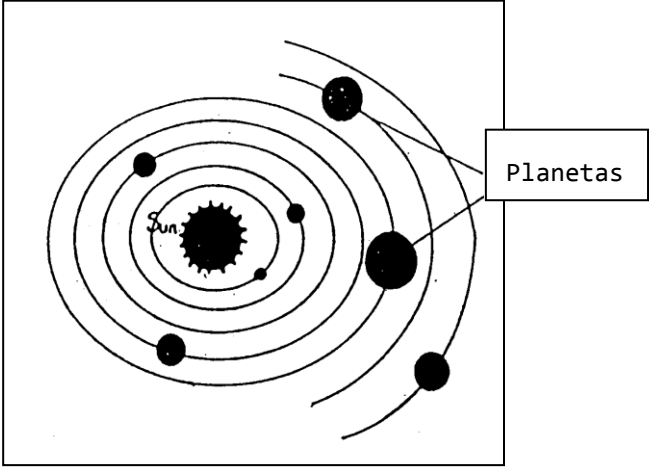
El ejemplo 1 (Rebote de pelotas) requiere que los estudiantes se identifiquen con la situación descrita así como la pregunta a abordar. Es probable que el tema sea familiar a todos los estudiantes, por lo que el nivel de conocimiento requerido es bajo y el desafío tiene que ver con conducir un test imparcial. El formato para responder, y el esquema de puntuación por cada respuesta correcta de cada cuadrado, baja las probabilidades de éxito basada en adivinanza. Esta es una manera de hacer una pregunta de opción múltiple en el que se reduce la posibilidad de éxito por adivinación. También significa que los estudiantes deben leer y entender las instrucciones para registrar su respuesta, de lo

76 DES, DENI and WO (1985) *APU Science in Schools Age 11 Report no 4*. London: HMSO

contrario se corre el riesgo de fracasar por otras razones que no están relacionadas con la habilidad requerida para dar respuesta a la pregunta.

Ejemplo 2: Los planetas (De APU 11 años Reporte N° 1)⁷⁷

Los planetas se mueven alrededor del sol.



Observa la siguiente tabla

Planeta	Distancia desde el Sol	Tiempo de viaje alrededor del Sol
Mercurio	58 millones de kilometros	88 días
Venus	108 millones de kilometros	225 días
Tierra	150 millones de kilometros	1 año
Jupiter	780 millones de kilometros	12 años
Uranio	2870 millones de kilometros	84 años
Neptuno	4500 millones de kilometros	165 años

a) Hay otro planeta que no está incluido en la tabla. Está cerca de 1430 millones de kilómetros desde el Sol. ¿Cuánto tiempo, piensas tú, tardará el planeta en hacer un viaje alrededor del Sol?

☐ 10 años

☐ 100 años

☐ 100 días

☐ 30 años

☐ 300 años

b) ¿Por qué crees tú que tomará ese tiempo?

Porque.....

.....

Comentario

En el Ejemplo 2 (Los Planetas) se da toda la información necesaria para responder la pregunta y los alumnos no necesitan saber nada acerca de los planetas. El foco de la tarea es que los estudiantes reconozcan el patrón entre las dos series de cifras dadas. Sin embargo, presentarla en "frío" a los estudiante (es decir, no como parte de un estudio más amplio sobre el sistema solar) podría parecer requiriendo más conocimiento del necesario. Algunos alumnos podrían no responderla si su primera impresión fuese que para hacerlo requieren saber acerca de los planetas. El puntaje se puede configurar de manera de dar algún crédito a la parte a) a pesar de que hay una posibilidad entre cinco

77 DES, DENI and WO (1981) *Science in Schools Age 11 Report No 1*. London: HMSO

de adivinar correctamente, pero mayor ponderación a la parte b) en función de si el patrón se indica o no explícitamente. La naturaleza abierta de la parte b) es más exigente que seleccionar entre razones alternativas determinadas. Proporciona más información acerca de cómo los estudiantes interpretan los datos pero es más complejo de calificar.

*Ejemplo 3: Cambio climático (De la evaluación PISA de Ciencias 2000)*⁷⁸

Lee la siguiente información y responde las preguntas que se presentan a continuación.

¿QUÉ ACTIVIDADES HUMANAS CONTRIBUYEN AL CAMBIO CLIMATICO?

La quema de carbón, petróleo y gas natural, así como la deforestación y las diversas prácticas agrícolas e industriales, están alterando la composición de la atmósfera y contribuyen al cambio climático. Estas actividades humanas han llevado al aumento de las concentraciones de partículas y gases de efecto invernadero en la atmósfera.

La importancia relativa de los principales contribuyentes al cambio de temperatura se muestra en la Figura 1.

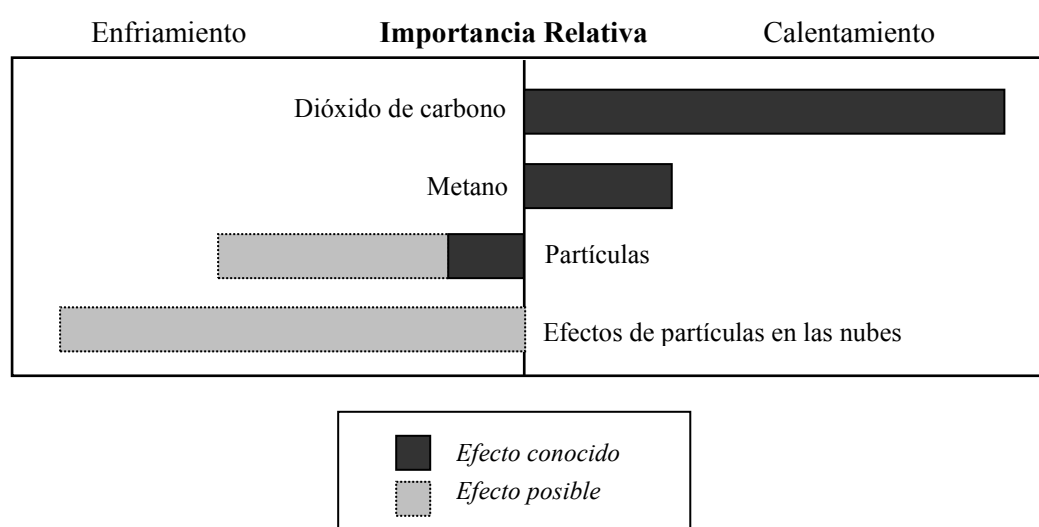


Figura 1: Importancia relativa de los principales contribuyentes al cambio de temperatura de la atmósfera. Fuente: adaptada de <http://www.gcrl.org/ipcc/qa/04.html>

Las barras que se extienden a la derecha de la línea central indican un efecto de calentamiento. Las barras que se extienden a la izquierda de la línea central indican un efecto de enfriamiento. El efecto relativo de las "Partículas" y los "Efectos de las partículas en las nubes" son bastante inciertos: en cada caso, el posible efecto está en algún lugar en el rango indicado por la barra de color gris claro.

La Figura 1 muestra que concentraciones elevadas de dióxido de carbono y metano tienen un efecto de calentamiento. Las concentraciones elevadas de partículas tienen un efecto de enfriamiento en dos formas, identificadas como "Partículas" y "Efectos de partículas sobre las nubes".

78 OECD (2000) op cit p 88

Item 1:

Utilice la información de la Figura 1 para apoyar la visión de que se debe dar prioridad a la reducción de las emisiones de dióxido de carbono de las actividades humanas mencionadas.

Item 2:

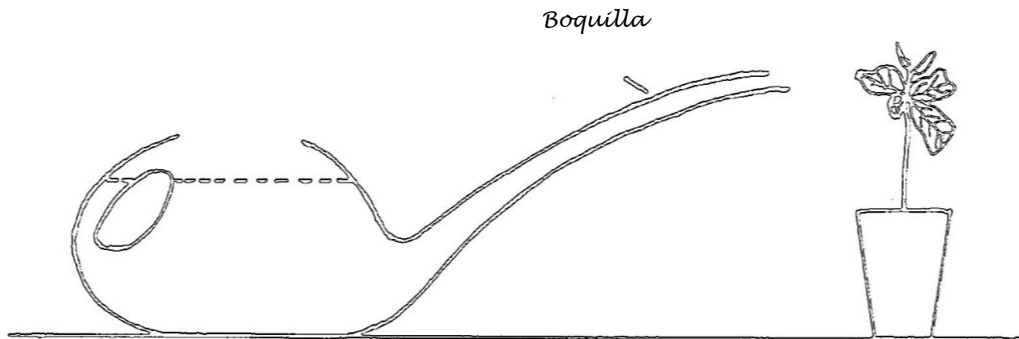
Utilice la información de la Figura 1 para apoyar la opinión de que los efectos de la actividad humana no constituyen un problema real.

Comentario

El Ejemplo 3 (Cambio climático) es otra pregunta en la que se da información y se les pide que la utilicen, en este caso para apoyar conclusiones alternativas acerca de la acción que los datos sugieren que podrían adoptarse. La información es auténtica y presenta la clase de problema con la que el alfabetizado científicamente debería ser capaz de involucrarse. Las dos partes de la tarea ilustran la incertidumbre de la interpretación de la información científica en ciertos casos. En teoría, se entrega toda la información y se les solicita a los estudiantes interpretar el gráfico. No necesitan saber cómo el dióxido de carbono, el metano, las partículas y sus efectos sobre las nubes causan el calentamiento y el enfriamiento. Sin embargo, sin ningún conocimiento sobre estas cosas, la pregunta es probable que carezca de sentido y es poco probable que se involucren con el problema planteado.

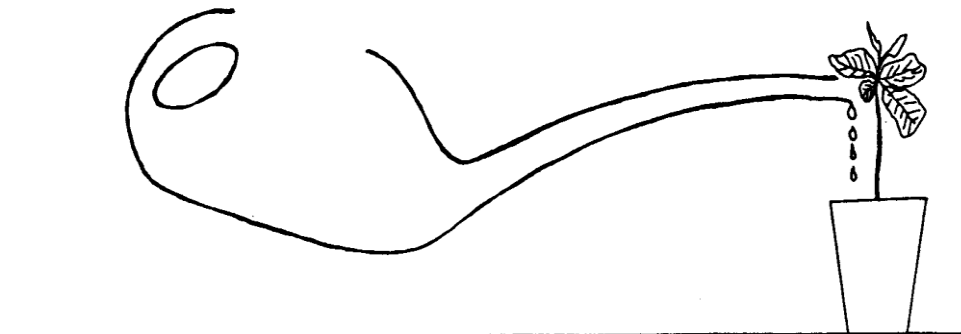
*Ejemplo 4: La regadera (De APU 11años, Reporte N°1)*⁷⁹

a) Las línea de puntos muestra donde está la superficie del agua en la regadera.



Dibuja una línea para mostrar dónde está la superficie del agua en la boquilla de la regadera.

b) La regadera está inclinada de modo que el agua comienza a gotear a través de la boquilla



Dibuja una línea para mostrar dónde está la superficie del agua ahora.

Comentario

El Ejemplo 4 (La regadera) requiere aplicación de conocimiento acerca del flujo del agua hasta alcanzar un nivel estable. Se asume que el contexto es familiar y la respuesta se solicita a través de un dibujo, por lo que es menos dependiente de la escritura y de vocabulario que las otras preguntas. Sin embargo, a pesar de que el esquema de puntaje permite inexactitud en el dibujo, el abordaje correcto de la tarea todavía requiere que los alumnos lean como se espera que la respondan.

⁷⁹ DES, DENI and WO (1981) op cit p 98

Comentarios generales sobre preguntas de pruebas escritas

Hay varios puntos que surgen de estos ejemplos. El más obvio es la demanda inevitable por la lectura y la comprensión de la pregunta y, dependiendo del formato de respuesta, la capacidad para escribir. A continuación, el intento de situar la tarea en un contexto que pueda parecer real para el estudiante significa que debe presentarse algún tipo de "historia" como contexto para la tarea. Los estudiantes tienen que leer e involucrarse con el contexto con el fin de responder a la pregunta. Tenemos que preguntarnos, entonces, si la naturaleza de este contexto hace una diferencia en cómo el estudiante responde. Por ejemplo, ¿sería la identificación de variables en el Ejemplo 1 más difícil para algunos estudiantes si el contexto fuera la comparación del efecto de cambiar los ingredientes en la fabricación de un pastel, o la velocidad de los autos de juguete pendiente abajo?, ¿se afectaría la capacidad de identificar un patrón en los datos si estos se presentan pictóricamente en lugar de números?, ¿se afecta el desempeño de los estudiantes según el formato de respuesta? La respuesta que ofrece la investigación a todas estas preguntas es que estas características sí importan y que el contexto de las preguntas es particularmente importante para la habilidad y disposición de los estudiantes para

Recuadro 17: Desigualdad de oportunidades en las pruebas

La investigación muestra que la familiaridad con el contexto en el que se plantean los ítems de las pruebas, es un factor que afecta el desempeño de los estudiantes. Cuando la familiaridad con el contenido está desigualmente distribuida en la población, se plantean interrogantes acerca de la equidad de la prueba. A algunos estudiantes se les puede estar negando la oportunidades de experiencias relevantes necesarias para comprender un ítem de la prueba, por ejemplo a las niñas o a los quienes viven en contextos desfavorecidos económicamente. Algunos estudiantes pueden tener dificultades con el idioma de la prueba⁷⁹ cuando esta puede ser su segunda o tercera lengua. Estos estudiantes y aquellos con necesidades físicas especiales o dificultades de aprendizaje, pueden no estar recibiendo las oportunidades de demostrar lo que pueden hacer y en consecuencia puede negárseles el acceso a continuar de su educación o formación.

participar y mostrar lo que pueden hacer. Otras circunstancias que plantean temas de equidad, especialmente para los estudiantes de los países con poblaciones diversas, se observan en el Recuadro 17.⁸⁰

Para reducir al mínimo el efecto de un contexto particular o formato específico, se debe incluir en una prueba una variedad de contextos y formatos. Puesto que existe un límite en la longitud de la prueba que se puede esperar que realice un estudiante, la consecuencia es que existe una tendencia a favorecer la inclusión de más preguntas cortas que pueden estar distribuidas en varios contextos, y más preguntas de respuesta cerrada que pueden ser respondidas mediante selección en lugar de ser respondidas con sus propias palabras, y que tienen la ventaja adicional de que pueden ser calificadas en forma automatizada.

Es importante notar que la restricción en longitud, y por lo tanto en la variedad de contextos, es más severa cuando se evalúa a estudiantes en forma individual con las mismas preguntas. No es un problema tan grave para el caso de un estudio poblacional ya que se puede utilizar un gran número de preguntas, divididas en un

80 Noble et al (2012) "I never thought of it as freezing": How Students Answer Questions on large-scale science tests and what they know about science, *Journal of Research in Science Teaching*, 49 (6) 778–803.

determinado número de sub-tests, cada uno aplicándose a una muestra aleatoria de la población. Este es el diseño utilizado en los estudios nacionales, como la NAEP y el APU y los estudios internacionales de TIMSS y PISA. El propósito de estos estudios es identificar el desempeño a nivel nacional o regional, que se deriva de la combinación de los resultados de los sub-tests. Los resultados individuales de los estudiantes, o incluso de toda una escuela, tienen poco valor en términos de los objetivos generales que se está evaluando, pero cuando se combinan a través de la muestra, el resultado proporciona una imagen mejor de los resultados nacionales comparado con los que se obtendría si todos los estudiantes tomaran las mismas preguntas de prueba. Cuando se dan diferentes preguntas a diferentes muestras de estudiantes, no hay necesidad de atiborrar una prueba corta con una cantidad de ítems breves y se puede otorgar más tiempo a los estudiantes para involucrarse con un determinado contexto. Por ello no es una coincidencia que los ejemplos aquí citados hayan sido creados y utilizados en los APU y PISA.

Recuadro 18: Fuentes de ítems y tareas para la evaluación en ciencias

SEAR (Science Education Assessment Resources/Recursos de Evaluación para Educación en Ciencias)⁸¹

Comprende un banco de recursos de evaluación que cubre todos los niveles de la educación escolar desde 1 a 12 años de edad. Las tareas están rotuladas y se puede acceder a ellas, de acuerdo al nivel de alfabetización científica, propósito evaluativo, tipo de tarea y el foco del objetivo de aprendizaje. Se definen seis niveles de alfabetización científica en el "Mapa de Progreso de Alfabetización Científica" que está relacionado con el marco de PISA. El banco incluye tareas para diagnóstico, propósitos formativos y sumativos y se presentan en distintos formatos como selección múltiple, preguntas abiertas y de actividad práctica. Un completo sistema de calificación acompaña a cada pregunta.

PISA (Programme for International Student Assessment/Programa Internacional de Evaluación de Estudiantes), conjunto de ítems de ciencias liberados⁸²

Esta es una colección de preguntas que han sido utilizadas en un estudio de alumnos de 15 años y publicadas en varios informes. Cada ítem incluye el desafío y el esquema de calificación.

APU (Assessment of Performance Unit/ Unidad de Evaluación de Desempeño), preguntas de ciencias.

En el informe escrito para docentes, se presentan ejemplos de preguntas escritas y ejercicios prácticos utilizados en estudios con alumnos de Inglaterra y Gales de 11, 13 y 15 años hecho en los años 1980.⁸³ Este incluye para cada pregunta, el esquema de calificación y la descripción de la respuesta dada por los estudiantes.

NAEP (National Assessment of Educational Progress – the Nation's Report Card/ Evaluación Nacional del Progreso Educativo – El Informe de la Nación) Muestra de Ítems de Ciencias 2011⁸⁴

Provee todas las preguntas liberadas de la evaluación de ciencias del 2011 aplicada a los niveles de 4°, 8° y 12° de educación. Los ítems incluyen formatos de selección múltiple, respuesta corta y de desarrollo. Para cada pregunta, hay una guía de calificación, respuesta de estudiantes y datos de desempeño.

81 Australian Government Department Education, Employment and Workplace Relations. *Science Education Assessment Resources (SEAR)* <http://cms.curriculum.edu.au/SEAR>

82 OECD (2006) PISA released items: Science. Paris: OECD <http://www.oecd.org/pisa/38709385.pdf>

83 <http://www.nationalstemcentre.org.uk/elibrary/collection/727/assessment-of-performance-unit-science-reports-for-teachers>

84 US Department of Education http://nationsreportcard.gov/science_2011/sample_quest.asp

En el Recuadro 18 se presenta una lista de fuentes donde se publican preguntas escritas e ítems de desempeño que se han utilizado en pruebas y estudios de ciencias. Estos incluyen, pero no están restringidos a, ítems que evalúan comprensión y aplicación de conocimiento y el uso de habilidades de indagación científica relevantes para evaluar el logro de las metas de ECBI.

Preguntas de desempeño

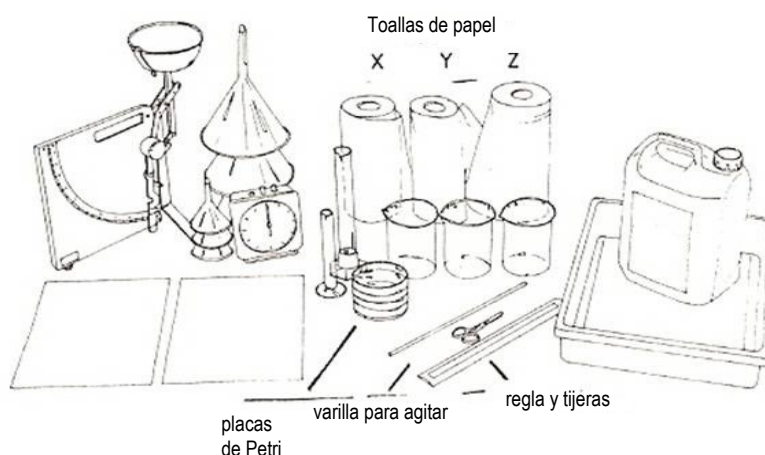
Dos de las deficiencias de las pruebas escritas - la dependencia de la lectura y de la escritura, y la necesidad de enfrentarse con un problema en contexto que se presenta sólo en el papel - se pueden evitar - en cierta medida - si el estudiante es capaz de llevar a cabo una indagación con objetos reales y con equipamiento de laboratorio. Hay, por supuesto, otras restricciones. La pregunta tiene que ser presentada al estudiante, que tiene que interactuar con ella como si se tratara de su propia pregunta, y la situación está lejos de ser aquella de una sala de clases normal, ya que los estudiantes pueden trabajar solos (a veces en parejas) con un administrador presente para observar sus acciones. Sin embargo, sí ofrece una oportunidad para que los estudiantes exploren, prueben cosas, y empiecen nuevamente si es necesario. El Ejemplo 5 muestra una investigación sobre el Papel Absorbente que se usó en un estudio con estudiantes de 11, 13 y 15 años de edad hecho por APU en Inglaterra en la década de 1980.

Ejemplo 5: Papel absorbente (Del Informe de Ciencias APU Para profesores no 6, 1985)⁸⁵

Tienes frente a ti, tres tipos de papel absorbente, marcados X, Y y Z.

Esto es lo que tienes que encontrar:

¿Qué tipo de papel es el que absorberá la mayor cantidad de agua?



Puedes utilizar todo lo que está frente a ti. Elige lo que creas es necesario para responder la pregunta.

Elabora un cuidadoso registro de tus resultados, de manera que otra persona pueda comprender lo que has encontrado.

⁸⁵ Welford, G., Harlen, W. and Schofield, B. (1985) *Practical Testing at ages 11, 13, and 15*. London: DES, WO and DEN

Comentario

Esta es una de las tareas administradas por un examinador capacitado, a estudiantes para que trabajen en forma individual. Los materiales son provistos por el programa de evaluación y son ordenados de igual forma para todos, previo a la llegada de los estudiantes a la sala. Se les presentan los materiales y se les da oportunidad para explorarlos. El examinador entrega a los estudiantes un papel donde está escrita la pregunta, con espacios para escribir notas de observación y hacer sus registros de resultados y se asegura que la pregunta haya sido comprendida. Durante la investigación, el examinador completa una lista de verificación, interviniendo sólo si hay un riesgo de seguridad. Finalmente cuando el estudiante ha terminado su ejercicio práctico, el examinador mira lo que el estudiante ha escrito para asegurar que se entiende y que ofrece evidencia para la respuesta. La prueba termina preguntando a los estudiantes qué cambio, si alguno, harían si tuvieran que comenzar nuevamente.

La reducción del error debido a algunos factores aumenta la influencia de otros, en particular el efecto del contexto. En una sesión de duración razonable, un estudiante cualquiera sólo puede realizar un número pequeño (dos o tres como máximo) de dichas investigaciones. Incluso en un estudio nacional, existe un límite para el número total que se puede utilizar y el efecto del contexto derrota cualquier intento de proporcionar un puntaje al "desempeño en las investigaciones" derivados del desempeño de varias investigaciones. Los estudios en que se han aplicado, entregan la evidencia que los estudiantes que se desempeñan bien en una investigación no necesariamente lo harán bien en otra que intenta evaluar las mismas habilidades en un contexto diferente.

Esto se confirmó en un estudio cuidadosamente diseñado en los EE.UU. Pine et al. (2006) evaluaron a los estudiantes de quinto grado usando varias tareas "prácticas" de desempeño, incluyendo una basada en "Papel Absorbente" y otra llamada "Resorte", sobre la longitud de un resorte cuando diferentes pesos se colgaban en él. No encontraron "esencialmente ninguna correlación para los puntajes individuales de los estudiantes". Los estudiantes con un 9 o 1 punto para la prueba de Resorte tenían para la prueba de Papel Absorbente puntajes que iban de 1 a 9.⁸⁶ Siendo que una tarea en particular es parte de una variedad más amplia de posibles tareas, el "error de muestreo de la tarea", es grande y significa que para obtener una calificación confiable de un estudiante en particular se requeriría que el individuo fuera capaz de enfrentar un número de tareas totalmente inaceptable. La única manera lógica de informar el desempeño es describiendo cada investigación y la variedad de formas diferentes en las que los alumnos responden.

86 Pine, J., Aschbacher, P., Rother, E., Jones, M., McPhee, C., Martin, C., Phelps, S., Kyle, T. and Foley, B. (2006) Fifth graders' science inquiry abilities: a comparative study of students in hands-on and textbook curricula, *Journal of Research in Science Teaching* 43 (5): 467-484. P 480

Alternativas a las pruebas

Del análisis que intenta proporcionar una evaluación confiable de la comprensión y las habilidades de investigación utilizando las pruebas, parece desprenderse que existen algunos obstáculos inevitables. Hemos señalado anteriormente, algunos requisitos claves para la evaluación de los conocimientos y habilidades de investigación científica. Estos eran

- que los estudiantes estén realmente involucrados en la realización de indagación
- que estén comprometidos con una pregunta que es nueva para ellos
- que el conocimiento que se requiere inevitablemente, pero no está siendo evaluado, esté disponible para ellos.

Cuando se presenta a los estudiantes una tarea en un contexto específico y restringido por un formato determinado, como es una prueba, surge un espacio de incertidumbre acerca de alcanzar los requerimientos. Una indagación científica genuina se lleva a cabo cuando los estudiantes tratan de responder una pregunta que es nueva para ellos y cuya respuesta no conocen anticipadamente. Pero, ¿quién ha de juzgar lo que es "nuevo" para un estudiante en particular?, ¿Puede la respuesta de un estudiante representar realmente el reflejo de su capacidad, si ésta ha sido creada al margen del contexto normal de aprendizaje y de la interacción con otros?

Una investigación realizada en Dinamarca⁸⁷ que usó preguntas de las pruebas PISA dio una respuesta clara a esta última pregunta. La investigación involucró solicitar a los estudiantes responder algunas preguntas PISA en forma oral por medio de una entrevista y participar en la conducción de una investigación realizada en parejas, descrita en una pregunta de la prueba PISA. La conclusión fue que "si se compara directamente y siguiendo los criterios de puntuación de PISA, los alumnos aumentaron en un 25% su desempeño cuando se les permitió aplicar sus conocimientos en un formato de prueba orientado socio-culturalmente".

Hay muchas otras preguntas y problemas que apuntan a las deficiencias de las pruebas, además de los problemas que surgen cuando se utilizan los resultados en la toma las decisiones y juicios como se señala en el Capítulo 4.

Aquí volvemos a la pregunta: ¿Existen alternativas a las prueba, para la evaluación sumativa? Afortunadamente sí existen, todas ellas dependiendo del hecho que las experiencias que los estudiantes necesitan para desarrollar las habilidades deseadas, la comprensión y las actitudes proporcionan oportunidades para evaluar su progreso. El factor clave es el juicio del profesor o profesora. La evaluación hecha por los profesores puede utilizar evidencia sobre las actividades regulares complementadas, en caso necesario, por la evidencia de tareas especialmente diseñadas para proporcionar oportunidades a los estudiantes de utilizar las habilidades y la comprensión que deben ser evaluadas.

Durante un período de tiempo – tal como un semestre o medio año – para el cual los logros se están informando, los estudiantes tienen la oportunidad de participar en una serie de actividades en las que puede desarrollar una serie de competencias. Estas mismas actividades ofrecen al profesor o profesora oportunidades para evaluar este desarrollo. En otras palabras, la limitación del rango de evidencia que puede obtenerse a través de una prueba no se aplica cuando la evaluación es realizada por el docente.

87 Dolin, J., & Krogh, L. B. (2010): The Relevance and Consequences of Pisa Science in a Danish Context. *International Journal of Science and Mathematics Education*, 8, 565-592.

Hay otras ventajas que van más allá de una evaluación válida de la comprensión y de las habilidades indagación científica, ya que este tipo de evaluación puede incluir una mayor gama de competencias. La observación durante el trabajo regular permite reunir información sobre los procesos de aprendizaje y no sólo sobre los productos. Incluso los tipos de pruebas que se muestran en el Ejemplo 5 no son capaces de evaluar cualidades como la reflexión sobre el proceso de aprendizaje. Esta información es útil para la selección de los estudiantes de cursos avanzados de formación profesional o estudios académicos, donde el grado en que los estudiantes han aprendido a aprender, y que está relacionado a la probabilidad de que se beneficien de un estudio más profundo, es tan importante como lo que ya han aprendido.

Implementación de la evaluación sumativa hecha por los profesores

La evaluación sumativa hecha por los profesores implica decidir, recopilar, interpretar y comunicar la evidencia para proporcionar un resumen de los logros de los estudiantes. Así como hay diferentes tipos de tareas y formas de responder a ellas en el caso de las pruebas, también hay diferentes enfoques de la evaluación de los profesores en función de cómo, cuándo y dónde se recoge la evidencia y cómo se interpreta y utiliza. Antes de analizar estos enfoques, deberíamos dejar en claro que la evaluación por los profesores no es sólo una cuestión de profesores utilizando sus juicios individuales acerca de qué evidencia usar y cómo interpretarla. La evaluación sumativa hecha por los profesores utiliza procedimientos acordados y está sujeta a medidas de control de calidad apropiadas para el uso de los resultados, es decir, recibe un control más estricto para su uso con alto impacto. El proceso requiere: algo de *evidencia o datos*, *criterios* para juzgarlo y *procedimientos* para llegar a un juicio.

La *evidencia o datos* relacionados con la comprensión de las ciencias y las habilidades de indagación científica pueden provenir de

- observar estudiantes que están involucrados en investigación científica
- un portafolio de los trabajos recogidos a lo largo del tiempo incluyendo recuentos, reflexiones, fotografías y otros productos de indagación
- los cuadernos de los estudiantes o publicaciones electrónicas
- presentaciones hechas por los estudiantes en forma individual o en grupos

Los criterios varían de acuerdo con el grado de especificación. Por ejemplo, un criterio altamente especificado sería "el estudiante sabe que los imanes atraen ciertos materiales y no otros, y se repelen entre sí". Esto sólo se puede aplicar al trabajo relacionado con los imanes. Mientras que "el alumno puede agrupar materiales según sus propiedades físicas" se puede aplicar a la información de una variedad de tareas de aprendizaje.

Un alto nivel de especificación de criterio significa que los profesores deben basar su evaluación en ciertas piezas de trabajo. Esto ocurre de alguna forma con la evaluación de portafolio, donde el tipo de tareas a incluir en el portafolio están prescritas y hay un criterio para cada una de ellas. Sin embargo, mientras mayor es la especificidad de la tarea, hay menos oportunidades de obtener evidencia de una amplia gama de actividades, que es la razón principal para usar la evaluación por profesores. Una revisión de la investigación sobre la validez y confiabilidad de la evaluación hecha por los docentes⁸⁸ concluyó que los enfoques más confiables eran aquellos donde los criterios eran detallados pero genéricos, siendo aplicables a una amplia variedad de actividades de sala de clases.

88 Harlen, W. (2004) Trusting teachers' judgements: research evidence of the reliability and validity of teachers' assessment for summative purposes, *Research Papers in Education*, 20(3); 245-270.

Los resultados de investigación indican que los criterios bien especificados ayudan al profesor a hacer juicios más confiables. El criterio más efectivo es el que guía la selección de evidencia sin prescribirla.

Los *procedimientos* para hacer juicios pueden comenzar a partir de los criterios, para luego, buscar la evidencia y contrastarla, o comenzar desde la evidencia y luego buscar en los criterios para ver cuál de ellos describe mejor el trabajo. En la mayoría de los casos existe una combinación de estos enfoques, con un ir y venir entre los datos y los criterios a fin de encontrar el "mejor ajuste". En lugar de tratar de asociar una determinada pieza de trabajo a un criterio particular, los profesores toman evidencias de varias actividades relevantes y generan un juicio "equilibrado", basado en el criterio que mejor se asocie a la evidencia. Los criterios también cumplen la función adicional de centrar la atención en los resultados de determinados tipos de trabajo, de manera que los docentes, estando alertados a buscar ciertos comportamientos particulares, tengan menos probabilidad de no evaluarlos.

Es común que los criterios sean identificados en diferentes "niveles", como se presenta en el Recuadro 19 (página 75), por lo que el resultado de la evaluación se puede expresar en términos del nivel en el cual el estudiante se está desempeñando actualmente. Los niveles son generados mapeando el progreso de los estudiantes en un área particular de aprendizaje, utilizando la evidencia de la investigación junto con la evidencia de la experiencia de los profesores y un poco de ensayo y error. Algunos ejemplos son los materiales desarrolladas en Australia⁸⁹ como "Developmental Assessment" (Evaluación para el Desarrollo) y las descripciones de desempeño en varios niveles que se establecen en una serie de estándares y metas nacionales de evaluación, como el Curriculum Nacional de Inglaterra (Recuadro 19). Sin embargo, la experiencia de uso (o, más exactamente, el mal uso) de los "niveles" ha dado lugar a una reacción en contra de su utilización para reportar el desempeño de los estudiantes. Los pros y los contras del uso de los niveles será discutido más adelante (página 80) ya que los argumentos se aplican igualmente a la utilización de pruebas. Pasamos ahora a algunas formas de mejorar la confiabilidad de la evaluación de los profesores.

Mejorando la confiabilidad de la evaluación hecha por profesores

La crítica más común expresada sobre la evaluación hecha por profesores se refiere a la confiabilidad de los resultados. En efecto, puede darse el caso que, cuando no se toman medidas para asegurar la calidad, los juicios de los docentes sean propensos a una serie de errores potenciales. La investigación ha reportado evidencia de sesgo al hacer un juicio de desempeño sobre una determinada tarea- una fuente no-aleatoria de error - debido a que los profesores, toman en cuenta información sobre aspectos no relevantes de la conducta de los estudiantes o por estar aparentemente influenciados por el género, las necesidades educativas especiales, o la capacidad general o verbal de un estudiante. Por ejemplo, se encontró que el comportamiento general influye en los juicios que profesores hacen de estudiantes pequeños sobre sus logros. Muy frecuentemente el comportamiento en el aula se vincula con el género en los primeros años de la escuela (las niñas pequeñas tienden a comportarse mejor que los niños pequeños), el resultado a veces se informa como un sesgo de género. El efecto es mucho menor para los estudiantes mayores. Otras causas de la baja confiabilidad se deriva de la inclusión de información irrelevante (como buena ortografía, cuando estos no son objetivos específicos de la tarea), la variación en la interpretación de los criterios, y el problema de relacionar el logro en contextos específicos con criterios necesariamente más generales. Sin embargo, hay varias maneras eficientes en las que la confiabilidad se puede mejorar, a un nivel igual e incluso superior a la de las pruebas. Los principales son: la moderación de grupo, la utilización de ejemplos, y el uso de prueba o tareas como verificación.

89 Masters, G. and Forster, M. (1996) *Progress Maps*. Camberwell, Victoria, Australia: ACER

Moderación de grupo

La moderación de grupo involucra la reunión de profesores para revisar las muestras de trabajo de los estudiantes, pero el propósito no es el de verificar las decisiones sobre trabajos particulares de los estudiantes, sino la de llegar a una comprensión compartida de los criterios y sobre cómo éstos se aplican. La intención es influir en el proceso de evaluación y de este modo garantizar una mayor confiabilidad de los resultados para todos los estudiantes. Los profesores llevan a la reunión varios ejemplos del trabajo de los estudiantes, ya que los juicios no deben hacerse en base a una única pieza de trabajo, describen el contexto del trabajo y discuten con otros como se hicieron los juicios. La moderación de grupo tiene beneficios más allá de la mejora de la calidad de los resultados de la evaluación. Ha quedado bien establecido que tiene funciones de desarrollo profesional. Las reuniones para discutir las inferencias que se pueden extraer de la observación de los estudiantes y del estudio de su trabajo proporciona a los profesores una mayor comprensión sobre el proceso de evaluación y mejora no sólo su evaluación sumativa, sino también el uso formativo de la evaluación.

Utilizando ejemplos

Proporcionar ejemplos de trabajos de los estudiantes y mostrar cómo ciertos aspectos se relacionan con los criterios de evaluación ayuda a transmitir lo que significan los criterios en la práctica. Los buenos ejemplos también indican las oportunidades que necesitan los estudiantes para demostrar sus logros en habilidades o comprensión. Aunque en los ejemplos se identifica el resultado de la evaluación, la atención se centra en el proceso para llegar a obtener ese resultado. Hay muchas fuentes de ejemplos de trabajos de los estudiantes o descripciones de sus acciones, anotados para resaltar características que son importantes en relación con los juicios que se hacen. Algunos de los materiales curriculares incluyen ejemplos de las acciones de los estudiantes, palabras, conversaciones, escritos o dibujos y discuten aspectos que conducen a una decisión acerca de si ciertos criterios claves se han cumplido. Por ejemplo, en cada Guía del Profesor del Programa de Ciencias a Nivel Primario de Nuffield (1995) se incluye un capítulo sobre la evaluación, donde el trabajo de los estudiantes se reproduce con un comentario sobre los aspectos que dieron lugar al juicio sobre el nivel alcanzado. El Ejemplo 6 se ha tomado de la Guía del Profesor sobre *Materiales* para los estudiantes de entre 8 y 11 años, donde un estudiante escribe sobre la investigación que su grupo hizo sobre la dureza de diferentes metales. (Esta pieza de trabajo es usada para ilustrar la situación, si bien – en la práctica – no debe usarse una única pieza de trabajo para evaluar el nivel de logro). (El ejemplo ha sido reproducido para su publicación en español).

Ejemplo 6: Informe de Matthew sobre la investigación de la dureza de los metales realizada por su grupo de estudiantes de 10 años.

Experimento

Doblando metales
 nosotros vamos a buscar que tan fuertes son algunos metales

Lo que hicimos
 Hicimos el experimento con diferentes metales tales como Níquel, Aluminio, latón, acero al carbono, acero inoxidable.

Los doblamos con un alicate hasta que se rompieran y aquí están los resultados.

Resultados

Níquel	0, muy difícil de romper
Aluminio	29 y se rompió
Zinc	13 y se rompió
Latón	57 y se rompió
acero al carbono	60 y se rompió
acero	56 y se rompió

Porque creemos que esto sucedió
 Encontramos que el acero al carbono fue el más fuerte y el más fácil de romper fue el Zinc, después fue el aluminio, no pudimos romper el Níquel porque era demasiado pequeño (un disco)

Los criterios aplicados en la evaluación (Recuadro 19) se obtuvieron a partir de los objetivos de logro nacionales ingleses (1995) para la indagación científica.

El comentario sobre el Ejemplo 6 explica:

La investigación del grupo de Matthew se derivó de su afirmación de que los metales eran duros. El profesor los animó a averiguar más sobre esto y a encontrar una manera de ver si los distintos metales eran igualmente duros. Su investigación fue comparar la facilidad con que ciertos metales se pueden romper cuando se doblan repetidamente hacia adelante y hacia atrás con un par de alicates. Ellos cuantificaron sus observaciones contando, pero no hicieron mención de la necesidad de controlar la cantidad de flexiones en cada caso, ni el tamaño de cada pieza de metal. Sin embargo, hicieron sus observaciones de manera sistemática, utilizando una tabla e interpretaron su hallazgo, lo que corresponde a un trabajo que se acerca al nivel 3. Se requiere poner más atención a la rigurosidad de sus pruebas para tener evidencia de logro en el nivel 3.⁹⁰

90 Nuffield Primary Science Teachers' Guide *Materials*. (1995) London: Collins Educational. p 86

Recuadro 19: Ejemplos de criterios de evaluación relacionados a la indagación científica

(Nota: un nivel abarca aproximadamente dos años, el nivel 2 siendo el nivel esperado para estudiantes de 7 años de edad)

Nivel 2: Hacer sugerencias así como también responder a las sugerencias de otros acerca de cómo descubrir cosas o comparar materiales. Utilizar equipamiento para hacer observaciones. Registrar lo que han encontrado y compararlo con lo que se esperaba.

Nivel 3: Decir lo que ellos esperan observar y sugerir formas de recoger información para someter a prueba su predicción. Llevar a cabo ensayos justos y saber por qué son justos. Registrar lo que han encontrado de variadas formas. Reconocer patrones en sus hallazgos.

Nivel 4: Hacer predicciones que guíen la planificación de su indagación. Utilizar equipamiento apropiado y hacer observaciones adecuadas y relevantes. Interpretar, sacar conclusiones e intentar relacionar hallazgos al conocimiento científico.

Nivel 5: Planificar investigaciones controladas sobre las predicciones que están basadas en conocimiento científico. Utilizar gráficos, mapas, tablas para registrar y ayudar a la interpretación. Considerar los hallazgos en relación al conocimiento científico.

Puesto que los profesores deberían basar sus juicios en una serie de trabajos del estudiante y no juzgar a partir de una sola pieza de trabajo, es muy útil disponer de material ejemplar en la forma de un portafolio de un estudiante más que diversas piezas individuales de trabajos de varios alumnos. Esto ayuda a los docentes a aplicar los criterios de una manera holística. Una mirada general a las descripciones de los niveles para las habilidades de investigación del Recuadro 19, por ejemplo, muestra que no todas las piezas de trabajo se ajustan a las descripciones y que tampoco estará cada uno de los criterios para un nivel representado en el portafolio. Lo que se recomienda es el "mejor ajuste", comparando los datos con el nivel más probable y con los inmediatamente por debajo y por encima de éste.

Los ejemplos pueden ser empleados en las reuniones de moderación de grupo, pero son particularmente útiles para profesores individuales que no pueden participar en estas reuniones.

Utilizando una prueba o tarea

En este enfoque, se utiliza una prueba corta común o tarea especial como medio de moderación o comprobación de los juicios de los profesores, pero no como una medida independiente de logro. Esta es la forma en que se usan las pruebas en Escocia para evaluar Inglés y matemáticas. Los profesores deciden sobre la base de una serie de evidencias proveniente de las actividades cotidianas realizadas durante un determinado tiempo, si un estudiante ha alcanzado un criterio particular a un cierto nivel en una materia. Una opción para moderar este juicio es que los profesores utilicen una prueba corta en el nivel que ha sido indicado por evaluación del profesor. La prueba, que los profesores administran y califican por sí mismos, se extrae de un banco externo de preguntas. El profesor compara los resultados de las pruebas con los resultados de la evaluación que realiza en sus propias clases. Esto se puede hacer siempre que el (la) profesor(a) juzga que un estudiante puede pasar la prueba, no en un momento fijo, y se administra en forma individual de manera informal lo que evita la ansiedad de los estudiantes. Cuando se confirma, a través de moderación, el nivel alcanzado según el juicio de los profesores se registra y luego se informa a su debido tiempo.

Ejemplos de evaluación hecha por profesores

La evaluación de estudiantes hecha por los docentes no es nueva; se utiliza comúnmente dentro de la entrega rutinaria de información por parte de las escuelas y para mantener registros cuando los resultados no son de alto impacto. En tales casos se requiere una moderación menos rigurosa que en los casos en que se toman importantes decisiones sobre la base de los resultados. En el Capítulo 3 se ha mencionado un ejemplo de evaluación por docentes que se utiliza para la evaluación con alto impacto de los estudiantes. Otros aspectos relevantes se consideran a continuación.

Evaluación hecha por profesores al final de la educación secundaria

El funcionamiento del sistema de evaluación escolar para el Certificación Superior en Queensland, Australia, es un ejemplo bien documentado de cómo este sistema puede funcionar. Ha existido desde 1972, cuando en Queensland se abolieron los exámenes externos. Las razones de preferir la evaluación realizada por los profesores refleja el argumento de que dicha evaluación es capaz de incluir una serie de logros de aprendizaje, tanto a nivel académico como vocacional de modo de apoyar más que de controlar el currículum. Adicionalmente, Maxwell sostiene que

Un principio importante de la evaluación hecha en la escuela es que la evaluación es progresiva y continua. Uno de los objetivos... es aliviar la alta presión de un examen final único – una sola prueba de la que dependa todo. Esto requiere no sólo que la evaluación se ajuste a la forma en que se implementa cada programa de la asignatura por la escuela, sino también que la evaluación ocurra progresivamente a lo largo de todo el curso de estudio. En otras palabras, la validez de la evaluación se mejora preparando un portafolio que contenga una variedad de tipos de evaluación y contextos. Así, también, la confiabilidad se mejora al tener muchas oportunidades para que el estudiante demuestre sus conocimientos y capacidades y mediante la recopilación de la información en diferentes ocasiones.⁹¹

El proceso se basa en el portafolio y permite la variación en el contenido de modo que las materias a incluir se pueden implementar con flexibilidad para cumplir con los requisitos locales. El elemento común es el sistema de criterios progresivos contra los que se juzga cada portafolio. También hay un fuerte sistema de moderación en el caso de aquellos contenidos que cuentan para acceder a la universidad, que asegura exitosamente la confianza de todos los interesados en el resultado de la evaluación.

El portafolio se construye a lo largo de los dos años del curso, durante los cuales su contenido va a cambiar no sólo a través de la adición de nuevo material, sino a través del reemplazo de la evidencia más antigua por evidencia más reciente. Es sólo la evidencia final la que se considera, aunque algunas se hayan obtenidos más tempranamente que otras. Así el portafolio final o de "salida" se compone de las tareas de evaluación que representan la información *más completa y más reciente* de los conocimientos y capacidades del alumno. Como explica Maxwell,

La *más completa* información significa que deberá proveerse información de evaluación para todos los aspectos obligatorios del plan de estudios. Los criterios importantes no se pueden soslayar; las evidencias de evaluación en el portafolio deben cubrir todos los aspectos requeridos del curso... La *última* información significa que las evaluaciones anteriores que ya no son relevantes pueden ser desechadas y reemplazadas por la evidencia más reciente... El objetivo final consiste en representar el estado del

91 Maxwell, G. (2004) op cit. p 2

conocimiento y la capacidad como típicamente demostrada por el alumno al final del curso.⁹²

Los criterios de evaluación, que se publican para que los estudiantes y padres de familia, así como los profesores, puedan estar familiarizados con ellos, describen lo que los estudiantes hacen en términos de descripciones de niveles. Por ejemplo, una de las subcategorías de "trabajar científicamente" se refiere a la planificación, y los criterios para ello se ordenan en cinco niveles o estándares de A hacia abajo en el Recuadro 20.

En el Recuadro 20, los criterios para los estándares A y B son los mismos, pero el juicio de "planificar" es sólo uno de los varios aspectos de "trabajar científicamente" que se juzgará como un todo. La comparación de la evidencia con los criterios implica juicios, no agregación, y es un juicio "equilibrado" del mejor ajuste.

La moderación implica varias etapas, comenzando con la aprobación del "plan de trabajo de la escuela" -los detalles de cómo la escuela tiene la intención de proporcionar oportunidades para que el

Recuadro 20: Criterios para evaluar portafolios (Queensland Senior Certificate)

Estándar A: planifica una serie de investigaciones científicas de problemas que incluyen muchos elementos nuevos y/o complejos

Estándar B: planifica una serie de investigaciones científicas de problemas que incluyen muchos elementos nuevos y/o complejos

Estándar C: planifica una serie de investigaciones científicas sobre problemas sencillos

Estándar D: participa en la planificación de algunas investigaciones científicas sobre problemas sencillos

Estándar E: participa en algunos aspectos de la planificación de investigaciones científicas de problemas sencillos

estudiante cumpla con los criterios finales para la evaluación de una materia. La moderación de los temas relacionados con la selección universitaria involucra paneles externos del distrito quienes revisan una muestra de portafolios de las escuelas y consideran la evidencia que apoya o cuestiona los "juicios" de la escuela. A su vez, un panel del estado revisa las muestras de los distritos y arbitra los casos difíciles.

La apertura de un proceso continuo de creación de un portafolio significa que al final del curso no debería haber sorpresas, ya sea para profesores o los estudiantes. Además, la "actualización selectiva" y la colección de la evidencia "más completa y más reciente" permite que los inicios

pobres, desempeños atípicos, y confusiones iniciales y temporales (por cualquier razón) sean ignoradas. Es importante destacar que estos procesos facilitan el uso de la evaluación para ayudar al aprendizaje, los estudiantes se benefician de la retroalimentación que reciben en evaluaciones tempranas. Ellos también tienen la oportunidad de auto-evaluación al tener que tomar la decisión de reemplazar en el portafolio una pieza de trabajo hecha tempranamente.

La experiencia de Queensland apoya el valor de la moderación colaborativa no sólo para garantizar una evaluación fiable, sino en cuanto provee desarrollo profesional.

La forma más poderosa para el desarrollo de competencias profesionales en evaluación es estableciendo conversaciones regulares entre profesionales sobre el desempeño de los estudiantes (conversaciones de moderación). Esto se focaliza mejor cuando se centra en ejemplos reales de portafolios de los estudiantes. Mirando los ejemplos reales y discutiendo las conclusiones que se pueden extraer sobre el desempeño de los estudiantes

92 Maxwell, G. (2004) op cit. p 4-5

en comparación con estándares explícitos agudiza el juicio de los profesores y construye conocimiento y experticia sobre la evaluación con más éxito que cualquier otro proceso.⁹³

Evaluación por los profesores al final de la escuela primaria en Inglaterra

Un ejemplo de la evaluación hecha por los docentes a los alumnos más jóvenes es la introducción en Inglaterra, en 1989, de la evaluación por docentes como parte de la evaluación nacional de los estudiantes a las edades de 7, 11, 14 y 16 años en Inglés, matemáticas y ciencias. Además de informar sobre su propia evaluación de los niveles de desempeño alcanzados por sus estudiantes, los profesores tenían que administrar pruebas diseñadas externamente. Cuando se introdujo⁹⁴ por primera vez, se reconoció que las pruebas externas podrían cubrir sólo una parte del plan de estudios, y la evaluación por los profesores se pensaba debía ser el componente principal. (Este reconocimiento no duró, por las razones indicadas en el Capítulo 4). El reporte formal de sus evaluaciones considerando los niveles del currículum nacional resultaba novedoso y atrajo la atención de los investigadores para averiguar cómo lo hacían los profesores de la escuela primaria. Gipps et al.⁹⁵ observaron y entrevistaron a profesores de alumnos de 11 años. Ellos encontraron varias aproximaciones diferentes, pero la mayoría tenían en común la obtención diaria de los datos mediante el uso de técnicas como pararse detrás, escuchar, hacer preguntas abiertas, observar y tomar notas, o bien usar su memoria. La mayoría de los profesores se preocupaba de perturbar la enseñanza normal tan poco como fuese posible. Algunos guardaban piezas significativas del trabajo de los estudiantes, mientras que otros mantenían un registro de las notas que habían dado a sus estudiantes. Si los profesores encontraban que no tenían suficientes datos de algunos estudiantes en relación con los objetivos de su trabajo, dirigían su atención hacia estos estudiantes. Al final del año, cuando se hacía necesario asignar un nivel al trabajo de cada estudiante, los profesores utilizaban el trabajo recogido y/o sus notas y recuerdos para decidir el nivel de logro con el enfoque "del mejor ajuste" en comparación con los criterios basados en niveles, similares a los que aparecen en el Recuadro 19.

Evaluación de las habilidades de indagación científica por profesores en escuelas básicas en Francia

Un cambio importante en el currículum fue introducido en Francia en el año 2006 en la forma de *Le socle commun de connaissances et compétences* (Base Común de Conocimientos y Habilidades). Esto establece los conocimientos que los estudiantes deben saber y las habilidades que deben haber adquirido al finalizar la enseñanza obligatoria (16 años) y le da a las escuelas la responsabilidad de lograr esto. Fue, y sigue siendo, un gran cambio en las responsabilidades de las escuelas. Para implementar la parte de este nuevo currículum relacionada con la ciencia, los profesores de básica de primer y segundo ciclo, se enfrentaron al gran desafío de enseñar y evaluar las competencias de indagación científica, lo que la mayoría no había hecho antes. El Ministerio proporcionó una libreta de registros para cada estudiante, con el propósito de que los profesores registraran al final de cada año si el estudiante ha alcanzado las habilidades específicas y conocimientos, con un simple "sí" o "no". Para proporcionar ayuda a los profesores con la implementación y la evaluación de las habilidades de indagación científica, el equipo del proyecto de *La main à la pâte* (LAMAP) ayudó a los grupos locales a desarrollar tareas que le permitiesen a los profesores evaluar las habilidades de indagación científica y los conocimientos de sus estudiantes. Las tareas incluyeron tanto preguntas cerradas como abiertas y algo de actividad práctica, como se aprecia en el Ejemplo 7.

93 Maxwell, G. (2004) op cit. p 7

94 DES/WO (1988) *National curriculum Task Group on Assessment and Testing: A Report*. London: HMSO

95 Gipps, C., McCallum, B. and Brown, M. (1996). Models of teacher assessment among primary school teachers in England., *The Curriculum Journal*, 7 (2) 167-183

Además, el Ministerio proporcionó una guía a los profesores para observar y cuestionar a los estudiantes y ayudarlos a evaluar a sus estudiantes tanto formativa como sumativamente.⁹⁶ Se proponen cuatro fuentes de datos para la evaluación de los aprendizajes: el estudio de los cuadernos de los alumnos; la observación durante las actividades de clase; los textos de los alumnos o relatos orales de una investigación experimental, incluyendo lo que hicieron y lo que encontraron, e ilustrado con diagramas; rendir una prueba estandarizada. Se da también orientación hacia que hay que considerar cuando se revisa el trabajo de los alumnos, cuando se les observa en sala y se escucha sus presentaciones.

Ejemplo 7: Una tarea para estudiantes de último año de educación básica en Francia de LAMAP⁹⁷

En esta tarea, los estudiantes primero leen una página de información sobre la capa de hielo alrededor de los polos norte y sur, y cómo ésta cambia del invierno al verano en ambas regiones. Ellos responden a algunas preguntas según la información proporcionada, incluyendo el efecto del alza de la temperatura sobre el hielo en los polos. Luego se les sugiere una investigación ilustrada por una fotografía que muestra dos vasos de agua. Un vaso muestra un cubo de hielo flotando en el agua, y el otro vaso muestra un cubo de hielo que se mantiene por encima de la superficie del agua. El nivel del agua en ambos vasos es el mismo. Se les pide a los estudiantes predecir lo que va a pasar con el hielo y el nivel del agua en los dos vasos después de un período de tiempo. Tienen que dar razones de sus respuestas. Se espera que los estudiantes lleven a cabo esta investigación en la práctica, pero también hay otra fotografía que muestra lo que sucedió después de que el hielo se derritió en caso de que esto no fuera posible. Una serie de preguntas piden a los estudiantes comunicar si lo que pasó estuvo de acuerdo con sus predicciones, así como evaluar la evidencia para una serie de posibles razones de las observaciones. Una última pregunta pide a los estudiantes comentar la declaración que se hace a menudo sobre el calentamiento global que "el derretimiento del hielo polar hará que el nivel del mar suba". Se solicita que respondan por escrito y en forma extendida utilizando un vocabulario adecuado.

Doble uso de los datos para la evaluación formativa y sumativa

En estos tres ejemplos se ve que hay oportunidades para el uso formativo de los datos recogidos para la evaluación sumativa. Este uso dual es más consciente y deliberado en el siguiente reporte (idealizado) de una serie de lecciones de ciencia en la que estudiantes de octavo grado (edad 13 años) estaban estudiando la transferencia de energía térmica a través de diferentes materiales.

Los estudiantes estaban investigando las propiedades de aislamiento de diferentes materiales que podrían utilizarse para hacer abrigo. Se les proporcionó envases de metal en los que podían poner agua para representar un "cuerpo" y pedazos de tela para envolver la parte exterior de los contenedores. Se colocaron termómetros en el agua para medir los cambios de temperatura del "cuerpo". Pero hubo que tomar varias decisiones en la preparación de las investigaciones. ¿Qué temperatura debería tener el agua que está dentro del recipiente? ¿Daría un resultado útil llevar a cabo la investigación en el laboratorio temperado en lugar de hacerlo al aire frío? ¿Cómo asegurarse de que la comparación fuese justa? Algunas de estas decisiones requirieron que los estudiantes aplicaran lo que sabían acerca de la conducción y otras formas de transferencia de energía térmica, mientras que otros alumnos requirieron comprender cómo hacer comparaciones justas.

96 http://cache.media.eduscol.education.fr/file/socle_commun/99/7/Socle-Grilles-de-reference-palier2_166997.pdf, page 36

97 <http://www.fondation-lamap.org/fr/page/14209/des-exemples-d-valuation-adapt-s-des-s-quences-d-enseignement>

Mientras estaban planeando lo que harían, la profesora les recordó la lista de cosas a tener en cuenta en la planificación de una investigación, una lista que había sido elaborada a partir de una discusión anterior y que estaba en sus carpetas. La profesora observó sus acciones y escuchó sus conversaciones, mientras ellos planificaron y realizaron sus investigaciones, ocasionalmente pidiendo a los grupos explicaciones de sus razones sobre cómo hacer ciertas cosas. En la etapa de presentación de informes sobre sus hallazgos hubo más oportunidad para que los profesores reunieran evidencia que podía utilizarse para ayudar a desarrollar la comprensión de los estudiantes sobre cómo la energía calórica se transfiere y así como sus habilidades de indagación.

De este modo, durante la lección, la profesora respondió a lo que vio y escuchó planteando preguntas para estimular que repensaran las decisiones; solicitando justificación de las afirmaciones o acciones, solicitando explicaciones sobre cómo ciertas partes de la investigación podría ayudarles a alcanzar su objetivo. En otras palabras, la profesora estaba usando la evidencia formativamente para ayudar al aprendizaje. Ella también tomó notas sobre la ayuda que algunos estudiantes habían requerido a fin de que esto pudiera ser objeto de seguimiento en futuras lecciones.

Luego, al final del año, cuando fue necesario informar el progreso en término de los niveles alcanzados, la profesora revisó la evidencia de ésta y otras lecciones de ciencias. Tanto para los objetivos conceptuales como para los objetivos de indagación científica, la evidencia de diferentes actividades específicas tuvo que reunirse para formar un juicio global sobre el logro de cada estudiante. Preparándose para esto la profesora dispuso tiempo para que los estudiantes revisaran sus carpetas, compararan sus reportes de investigaciones más actualizadas con los iniciales en términos de cómo la evidencia había sido recogida y utilizada para llegar a un resultado y a una explicación del resultado. Luego ellos seleccionaron los mejores ejemplos del trabajo con estas características. Al proporcionar tiempo de clase para esto, la profesora fue capaz de hablar en forma individual y así garantizar que la selección de criterios fuese entendida y aplicada apropiadamente. Luego ella contrastó la evidencia con los criterios de información sobre los niveles de logro, en este caso las descripciones para los niveles 4, 5 y 6 de la "indagación científica" y de los "procesos físicos" del currículo nacional Inglés. Los resultados fueron revisados por el jefe del departamento y la evidencia de una muestra de tres estudiantes se debatió en una reunión de moderación a nivel departamental.⁹⁸

Informando la evaluación sumativa: pros y contras de utilizar niveles

Los niveles, como se señaló anteriormente, establecen un estándar (nacional) de comparación de acuerdo al cual se juzga el logro de los alumnos. Los estándares encarnan una visión del progreso en el aprendizaje de los estudiantes en términos descriptivos, pero también pueden convertirse en una escala, en el que los niveles se definen en términos de puntos en el progreso global. Los niveles son, por lo tanto, una forma abreviada de señalar el logro de aprendizaje y tienen la ventaja de permitir el análisis de los resultados de grupos de estudiantes. Los resultados de las pruebas también se pueden convertir en niveles mediante la asignación de niveles a rangos particulares de puntajes.

Problemas con los niveles

Sin embargo, el uso de niveles ha sido cuestionado en el transcurso de la revisión del currículo y la evaluación en Inglaterra por razones que sin duda tienen relevancia en otros sistemas donde el número de estudiantes que alcanzan ciertos niveles se ha convertido en una medición de alto impacto

98 Harlen, W. (2007) Op cit p129

para los profesores y la escuela. El informe de un influyente grupo de expertos en 2012⁹⁹ hizo algunas fuertes críticas al uso de niveles para informar del progreso de los estudiantes, sobre todo en la escuela básica, aceptando que los estudiantes son inevitablemente diferenciados en los últimos años de la enseñanza media. Los principales puntos que pueden resonar en otros sistemas de evaluación son los siguientes:

- Otorgar "niveles" fomenta un proceso de diferenciación de los alumnos llegando al extremo que los estudiantes se etiquetan en estos términos.
- Algunos estudiantes se preocupan más por saber en "qué nivel están" que por la sustancia de qué es lo que saben, pueden hacer y comprenden.
- La asignación de niveles aumenta la diferenciación social en lugar de llamar a esforzarse por asegurar el aprendizaje para todos los estudiantes.
- Describir a un estudiante como habiendo alcanzado un cierto nivel no transmite nada sobre lo que esto significa en términos de lo que el estudiante puede hacer, ni tampoco indica lo que es necesario para avanzar.
- Cuando los niveles se aprecian como importantes, los profesores, los padres y los estudiantes los utilizan de manera indebida para etiquetar a los estudiantes.
- Los estudiantes que se consideran como incapaces de alcanzar los niveles de logro tienen escasas oportunidades de progresar, aumentando la brecha de desempeño entre los estudiantes con altos niveles de logro y bajos niveles de logro.
- Puesto que los niveles, en general, están bien espaciados (separados por dos años) la tendencia ha sido la de crear sub-niveles, con el fin de mostrar el progreso. Sin embargo, estos tienen poca evidencia de base en el desarrollo cognitivo y sólo sirven para prescribir el plan de estudios más estrechamente.

En general, entonces, se ha encontrado que informar el desempeño de los estudiantes, ya sea evaluado por pruebas o por el juicio de los profesores, como un "nivel" de logro tiene un profundo impacto en cómo los profesores, los padres y los propios estudiantes, juzgan el progreso, con implicaciones para la motivación de los alumnos y el aprendizaje. Estas son consecuencias involuntarias de un marco excesivamente prescriptivo para el currículo y la evaluación.

Alternativas a los niveles

Una alternativa a reportar por niveles, es identificar los resultados de aprendizaje que se deben alcanzar al final de las etapas fundamentales de la educación. Las etapas claves pueden abarcar dos o tres años, por lo que los años entre los grados 1 a 12 podrían dividirse en cinco o seis etapas. Los resultados de aprendizaje esperados podrían ser similares a lo señalado en los mapas de progreso del currículo, pero focalizando en unas pocas competencias claves. Los juicios podrían ser sobre el logro de estas competencias claves con el objetivo explícito de garantizar que todos los estudiantes las hayan logrado al final de la etapa clave. El juicio de los profesores durante la etapa clave sería identificar a qué distancia están los estudiantes de alcanzar los resultados esperados o donde necesitan atención especial. Así pues, la evaluación cambiaría de buscar el "mejor ajuste" en la escala de niveles hacia el seguimiento del logro de las competencias claves.

La experiencia de los países con altos niveles de logro y pequeñas brechas entre los alumnos con mayores y menores logros, no indica que este enfoque produciría mayor repitencia. Lo que caracteriza a los sistemas de alto logro parece ser su enfoque a la progresión de los estudiantes y la diferenciación. En lugar de una cruda categorización del logro de los estudiantes, se estimula a todos los alumnos a alcanzar una comprensión adecuada antes de pasar al siguiente tópico o área. No hay

99 Department for Education, (2011). *The Framework for the National Curriculum. A report by the Expert Panel for the National Curriculum review*. (London: Department for Education).

ninguna suposición de que la habilidad limita el logro, lo que podría tener un efecto negativo en las expectativas de los estudiantes. Más bien, se supone que el compromiso profundo y el esfuerzo conducen a la comprensión y el logro para todos.

Ventajas y desventajas de las pruebas y de la evaluación hechas por profesores

En resumen, reuniremos algunas de las ventajas y desventajas del uso de pruebas o la evaluación hecha por docente con fines sumativos. Los principales factores tiene que ver con el equilibrio entre la validez y fiabilidad (véase el Capítulo 1 p10). Los tests proporcionan información sólo sobre una muestra de los objetivos, pero se pueden tomar medidas para garantizar una fiabilidad óptima. La lucha por la alta confiabilidad, sin embargo, conduce a la inclusión de objetivos que se pueden evaluar más fácilmente en forma confiable, lo que tiende a excluir objetivos que son menos fáciles de evaluar, como las habilidades superiores y las prácticas. En el caso de la evaluación hecha por los docentes, se puede incluir una gama más amplia de logros y de resultados de aprendizaje, pero la fiabilidad puede ser baja, a menos que se tomen medidas para garantizar la aplicación de estándares comparables. Sin embargo, los procedimientos de moderación proveen un desarrollo profesional valioso para los profesores.

Otros factores a tener en cuenta para decidir el balance adecuado en un caso particular son:

- Las pruebas proveen a los profesores con ejemplos claros del significado de los objetivos de aprendizaje, pero dirigen la enseñanza en direcciones específicas que son iguales para todos los estudiantes. La evaluación por los profesores les entrega mayor libertad para abordar los objetivos de aprendizaje utilizando enfoques que se adapten a sus estudiantes.
- Las pruebas que se proporcionan externamente a la escuela permiten a los profesores distinguir entre el rol de profesor y el rol de evaluador. La responsabilidad sobre la evaluación unifica el rol de profesor y de evaluador y puede parecer incrementando la carga de trabajo de los profesores.
- En algunas circunstancias, los resultados de las pruebas pueden utilizarse para proporcionar una retroalimentación que ayuda al aprendizaje, pero en general estas oportunidades para el uso formativo son limitadas. Cuando los profesores reúnen las evidencias del trabajo en curso de los estudiantes, ésta puede ser utilizada formativamente, para facilitar el aprendizaje, así como para fines sumativos.
- Es bien sabido que las pruebas inducen ansiedad en muchos estudiantes. No todos se afectan por igual y el impacto es a menudo exacerbado por la aplicación frecuente de pruebas. La evaluación hecha por los profesores reduce esta fuente de error y la desigualdad.
- El tiempo dedicado por los profesores a preparar a los estudiantes para las pruebas puede utilizarse más eficazmente para el aprendizaje cuando la evaluación es continua. Los recursos financieros también se liberan cuando se compran menos pruebas comerciales.

- Los usuarios de datos de evaluación tienen a menudo más confianza en las pruebas que en las evaluaciones hechas por los profesores, especialmente para los estudiantes mayores. Cualquier cambio requiere mayor apertura acerca de mejorar la precisión de las pruebas y los procedimientos de la evaluación realizada por los profesores.

Capítulo 7

Cambiando las prácticas de evaluación

La interacción entre la evaluación, el contenido curricular y la pedagogía representada en la Figura 3 (página 30) indica que el cambio en evaluación involucrará cambios amplios en la práctica. No es solo una cuestión de sumar nuevos procedimientos a una práctica de aula que aparte de eso no haya cambiado las interacciones y los procesos. Este es particularmente el caso cuando los cambios en evaluación requieren de un mayor uso de los juicios de los profesores, requiriendo habilidades y comprensión que son distintas de aquellas empleadas en la enseñanza y evaluación tradicional. Aunque existe todavía poca experiencia de cambio en la evaluación de ECBI, más allá de la escala de los ensayos piloto, hay mucho que aprender sobre cómo se ha provocado el cambio en otros aspectos de la educación. Así, en este capítulo, en primer lugar, examinaremos brevemente las lecciones que se pueden aprender de los intentos de realizar cambios en el contenido del currículo y en la pedagogía. Luego una revisión de algunos ejemplos de procesos de cambios en evaluación, conduce a realizar algunas sugerencias sobre lo que sería necesario hacer para generar un cambio en la evaluación de ECBI. Mientras que la experiencia descrita aquí ha sido en el contexto del desarrollo profesional de los docentes en servicio, los enfoques y los principios son igualmente relevantes para la formación docente inicial.

Aproximaciones hacia el cambio de las prácticas en educación

Las diferentes aproximaciones para cambiar las prácticas en el ámbito de la educación se pueden dividir en dos grupos principales: de transmisión y de transformación¹⁰⁰. Varias formas de *transmisión* implican la distribución de recursos que presentan ideas y ejemplos de nuevos contenidos y prácticas después de su desarrollo y publicación. Los recursos producidos son a menudo guías escritas asociadas con materiales audio-visuales. Cuando se han creado asociados a la implementación de nuevos planes de estudio, contienen implícita la promesa de ser una buena solución para producir los cambios. Este enfoque parece decir que todo lo que se requiere para implementar el cambio es que los profesores sigan las guías, ya que el pensamiento necesario para introducir las innovaciones lo habrían hecho otros por ellos. Sin embargo, este enfoque de "arriba hacia abajo", ha perdido reconocimiento al comprobarse que los mensajes recibidos en el aula, que intentan influir en las acciones que ahí se ejecutan, rara vez coinciden con las intenciones de los desarrolladores del material.

La evidente necesidad de una mediación de los mensajes a través de sesiones de desarrollo profesional lleva a la aproximación de "cascada". Esto comienza con un grupo que ha sido entrenado en el tema y en las prácticas esperadas y algunas veces en la forma de entrenar a otros grupos. Estas personas entrenadas, luego entrenan a otros y así la formación pasa de un grupo a otro y, finalmente, a los profesionales en el aula. Este enfoque tiene desventajas bien conocidas: hay mucho potencial para la distorsión de los mensajes en cada etapa de la cascada. "Pilotear y ampliar" es otra de las aproximaciones con algunas similitudes, pero comenzando con ensayos y la modificación de las ideas en unas pocas escuelas para garantizar la viabilidad y proporcionar credibilidad entre los docentes.

100 Hayward, L. (2010) Moving beyond the classroom, in J. Gardner et al *Developing Teacher Assessment*. Maidenhead, UK: Open University Press.

El éxito limitado de los enfoques de transmisión ha llevado a las alternativas descritas como *transformacionales*. La experiencia en muchas áreas de cambio en la educación, ya sea en el currículum, la pedagogía, la evaluación o en la organización y gestión escolar, es que la forma más efectiva de estimular el compromiso con el cambio es asegurar la participación en el desarrollo de nuevos procedimientos o materiales de aquellos que deberán ponerlos en práctica. Cuando los grupos de profesores trabajan junto a investigadores o desarrolladores, tiene la oportunidad de ser creativos y experimentales en un ambiente seguro, aprender de las ideas de los otros, combinar ideas y lograr la apropiación de las prácticas emergentes en lo que se describe como una aproximación "desde abajo hacia arriba". Ofrecer oportunidades para reflexionar y desarrollar la comprensión de los principios que subyacen a los cambios, y la experiencia, puede ser la forma más efectiva de aprendizaje profesional. Pero también es muy demandante de recursos, particularmente de tiempo, y claramente no puede expandirse a un gran número de profesores. Sin embargo, indica el tipo de experiencias que pueden ser efectivas en casos donde el cambio requiere más que adoptar nuevas técnicas.

Los enfoques *transformacionales* reconocen que el cambio en la práctica es una cuestión de aprendizaje y que el aprendizaje efectivo por los profesores tiene las mismas cualidades que para los estudiantes. Así como los estudiantes desarrollan la comprensión a través de su propia actividad mental y física, los profesores aprenden mejor cuando participan activamente en la transformación de su práctica. Hayward¹⁰¹ cita a Bruner¹⁰² al argumentar a favor de cuatro componentes esenciales del aprendizaje efectivo:

- agenciamiento, las personas toman un mayor control de su propia actividad de aprendizaje
- reflexión, dando sentido a lo que se está aprendiendo, entendiéndolo, interiorizándolo
- colaboración, compartiendo los recursos de la "diversidad de los seres humanos" involucrados en la enseñanza y el aprendizaje
- cultura, la construcción por individuos y grupos, de formas de vida y de pensamiento particulares que ellos llaman realidad.

Otro factor, revelado a partir de la evidencia de tres grandes proyectos de desarrollo en la escuela secundaria de Dinamarca¹⁰³, es el grado en que los profesores consideran una innovación como coincidente con sus (a menudo tácitas) creencias y valores en la educación. Cuando hay un alto grado de acuerdo los docentes son más propensos a aceptar nuevas ideas y hacer los cambios necesarios en sus prácticas. Si hay algún conflicto con sus valores, la participación de los docentes bien pueden carecer de compromiso real y el proyecto tener poca influencia sobre sus prácticas.

En las aproximaciones transformacionales, en lugar de asumir que todas las salas de clases son las mismas - y por lo tanto que una misma solución se ajustará a todas - la suposición es que los diferentes entornos de aprendizaje requerirán soluciones diferentes. Por supuesto, el problema de llegar a un gran número de profesores se mantiene. Las oportunidades para involucrar a los docentes en el desarrollo genuino y de adaptar las experiencias a las necesidades individuales a gran escala, son ciertamente limitadas. Sin embargo, es posible favorecer situaciones en las que los profesores puedan aprender en colaboración y encontrar la manera de poner en práctica las nuevas ideas y

101 Hayward, L (2010) op cit p 96

102 Bruner, J. (1996) *The Culture of Education*. Cambridge, MA: Harvard University Press

103 Dolin, J., Laursen, E., Raae, P. H., Senger, U. (2005). *Udviklingsprojekter som læringsrum. Potentialer og barrierer for skoleudvikling i det almene gymnasium*. Gymnasiepædagogik nr. 54, Syddansk Universitet. 232 s. (Development projects as learning arenas. Potentials and barriers for school development in the general upper secondary school. University of Southern Denmark)

alcanzar nuevas metas en el contexto particular de sus propias salas de clases, como se verá en algunos de los ejemplos más adelante. Se ha encontrado que las oportunidades para que los profesores visiten otras salas de clases o para mostrar videos de su enseñanza contribuyen al mejoramiento de sus aprendizajes, mientras que proyectos como Fibonacci¹⁰⁴ y Pollen¹⁰⁵ han demostrado lo que se puede hacer cuando el tiempo y los recursos se utilizan para que los profesores y formadores de profesores aprendan unos de otros.

Algunos ejemplos de cambio en las prácticas de evaluación

Tanto las prácticas de uso de la evaluación para ayudar en el aprendizaje y la evaluación sumativa en las que los profesores toman un papel de liderazgo requieren estrategias que difieren considerablemente de aquellas que son familiares para muchos profesores. Los primeros trabajos realizados en el cambio de las prácticas de evaluación centradas en la evaluación formativa, una vez que ésta se conoció ampliamente a través de la obra de Black y Wiliam¹⁰⁶, mostraron que implementar aspectos de la evaluación formativa podía elevar significativamente el nivel de logro de los estudiantes (ver Capítulo 3). Para muchos, se trataba de un uso nuevo de la evaluación y uno que recibía evidentes respuestas positivas de los estudiantes. Al mismo tiempo, se notó que las aproximaciones existentes hacia la evaluación sumativa no proporcionaban información coherente con los objetivos de una educación moderna y que para mejorar el grado de consistencia se requería de una mayor participación de los docentes en el proceso.

Profesores desarrollando prácticas de evaluación formativa: la experiencia en Inglaterra

El Proyecto de Evaluación Formativa King's Medway Oxford (KMOFAP) fue el precursor de varios proyectos que desarrollo del uso formativo de la evaluación¹⁰⁷. KMOFAP comenzó poco después de que Black y Wiliam completaron su revisión de investigación en el aula y reconocieron la importancia de encontrar formas en que los profesores pudiesen incorporar la evaluación formativa en su trabajo. Investigadores universitarios del King's College y personal asesor de dos autoridades locales colaboraron en la planificación del trabajo con dos profesores de matemáticas y dos profesores de ciencias de cada una de las seis escuelas secundarias. El financiamiento les permitió a los profesores ser liberados para sesiones de desarrollo profesional que constaban de siete sesiones de día completo repartidas en 18 meses. Después de introducir a los profesores y a los asesores a los aspectos de la práctica que la investigación mostraba eran efectivos para mejorar el aprendizaje, las primeras sesiones tuvieron como objetivo que los profesores elaboraran planes de acción para su implementación en el año escolar siguiente. Esto dio a los profesores tiempo para "experimentar con algunas de las estrategias y las técnicas sugeridas por la investigación, como cuestionamiento profundo, evaluación sólo a través de comentarios, compartir criterios con los estudiantes, y la auto-evaluación y evaluación por pares de los estudiantes."¹⁰⁸

Durante el proyecto, los miembros del equipo visitaron las salas de clases de los profesores para observar y discutir lo que estaba sucediendo y cómo se relacionaba con los planes de acción. Estas visitas fueron descritas como "no directivas, sino más bien como sostener un espejo para los

104 www.fibonacci-project.eu

105 www.pollen-europa.net

106 Black, P. and Wiliam, D. (1998) Assessment and classroom learning, *Assessment in Education*, 5 (1) 7-74

107 This account draws on Harlen, W. (2010) On the relationship between assessment for formative and summative purposes, in Gardner, J. et al *Developing Teacher Assessment* Maidenhead, England: Open University Press pages 100 – 129.

108 Wiliam, D. Lee, C., Harrison, C. and Black, P. (2004) Teachers developing assessment for learning: impact on student achievement, *Assessment in Education*, 11 (1) 49-66, p 54.

profesores".¹⁰⁹ Los profesores fueron introducidos a estrategias generales y a alguna idea de lo que se puede aspirar, pero sin dar modelos de lo que debían hacer. En consecuencia, el aporte de los investigadores fue diseñado para involucrar a los docentes en la identificación de cómo poner en práctica varias características de la evaluación formativa. Este rol de los investigadores resultaba posiblemente inusual y poco confortable para los profesores que los consideraban expertos. Los investigadores notaron que:

Al principio, daba la impresión de que los profesores no creían esto. Parecían creer que los investigadores estaban operando con un modelo perverso de aprendizaje por descubrimiento en el que los investigadores sabían muy bien lo que querían que los profesores hicieran, pero no se lo decían, porque querían que los profesores "lo descubrieran por sí mismos". Sin embargo, después de un tiempo, se hizo evidente que no había un modelo prescrito de acciones efectivas en la sala de clases, y que cada docente tendría que encontrar su propia manera de aplicar estos principios generales en sus propias aulas.¹¹⁰

Mientras los profesores probaban las actividades en la práctica, se dieron cuenta de la necesidad de comprender por qué estas actividades son importantes y por qué "funcionan". Mientras los docentes aplicaban las actividades y observaban cómo los alumnos reaccionaban a ellas, los docentes querían saber más acerca de la manera de como los estudiantes aprenden. Por lo tanto, aproximadamente un año después de comenzar el trabajo centrado en las acciones de sala de clase, una de las sesiones en servicio fue diseñada para introducir a los profesores a las teorías de aprendizaje y explicar la importancia de que los estudiantes sean activos en su aprendizaje.

En relación a las estrategias para la transformar los resultados de investigación en prácticas de sala de clases, los investigadores informaron que las nuevas prácticas desarrolladas por los profesores eran

mucho más ricas y más amplias que las que habíamos sido capaces de sugerir al principio del proyecto sobre la base de la literatura de investigación. Todas ellas implican (implicaban) cambio en la forma en que ellos trabajan con sus alumnos y con el curriculum.¹¹¹

En términos de cambio en los profesores, hubo diferencias esperadas en la respuesta entre los 24 profesores. Sin embargo, los investigadores informaron cambios en varios aspectos, en todos los profesores implicados en el proyecto. En particular, ellos observaron cambio en:

- la forma en que los docentes enfocaban su objetivo a ayudar a los estudiantes a aprender, a diferencia de "pasar el currículum"
- las expectativas que los profesores tienen de sus estudiantes, en el sentido que todos tenían la capacidad de aprender en un determinado momento y con la aproximación correcta
- los docentes le dan a los estudiantes un mayor control y comparten con ellos la responsabilidad de aprender

Black y sus colegas afirmaron que los cambios en la práctica tardaron en aparecer, pero fueron duraderos, y que probablemente no habrían ocurrido si los investigadores hubiesen suministrado recetas para lograr lecciones exitosas. Ellos consideraron que, "el cambio en las creencias y valores son el resultado de que los profesores se vieron a sí mismos como aprendices y trabajando con nosotros para aprender más"¹¹². Wiliam et al¹¹³, usando una variedad de comparaciones de clases, exámenes

109 ibid, p 54

110 Ibid, p 51

111 Black, P et al (2003) op cit p 57

112 Ibid, p 98.

113 Wiliam, D. et al (2004) op cit

escolares internos y externos y los resultados de las pruebas del currículum nacional, reportó un impacto positivo en los logros de los alumnos, incluso después de un año de intervención.

Escalando el Desarrollo Profesional en la Evaluación Formativa: la experiencia de USA

Desde 2003 al 2006 Wiliam, trabajando en ETS en los EE.UU. con Leahy, un director de escuela jubilado de Inglaterra, experimentó con formas de lograr los mismos efectos que se encuentran en el proyecto KMOFAP a una escala necesaria para llegar a un gran número de salas de clases. Ellos reconocieron que

cualquier modelo de desarrollo profesional docente efectivo y escalable, tendría que lograr un delicado equilibrio entre dos exigencias contradictorias. La primera era la necesidad de asegurar que el modelo era lo suficientemente flexible para permitir que se adaptara a las circunstancias locales de la intervención, no sólo para permitir que tuviese éxito, pero también para que pudiera capitalizar sobre cualquier situación presente del contexto local mejorando la intervención. El segundo era asegurar que el modelo fuera lo suficientemente rígido para garantizar que las modificaciones que tuvieran lugar preservaran suficiente fidelidad al diseño original para proporcionar una seguridad razonable de que la intervención no experimentaría una "mutación letal".¹¹⁴

Los investigadores desarrollaron y pilotearon una serie de modelos para trabajar con los profesores de los que obtenían conclusiones sobre el ritmo, la longitud, la duración y el contenido de sus intervenciones. Por ejemplo, encontraron que los intervalos de dos semanas entre las reuniones dieron muy poco tiempo a los profesores para planificar e implementar los cambios y tener algo que informar en la próxima reunión. Se encontró que las reuniones mensuales fueron lo óptimo. Las reuniones de dos horas de duración resultaron ser demasiado largas en algunos casos, pero una hora de reunión demasiado corta, por lo que se necesitaba algo intermedio. En términos del número de participantes, entre 8 y 12 se encontró que era lo mejor, con el grupo incluyendo a los docentes de un rango de especialidades. Se encontró también que el progreso fue ayudado por la adopción de un programa estructurado para cada reunión, entonces los profesores sabían qué esperar y cuál era su rol en el proceso.

Las ideas que surgieron de estos ensayos se utilizaron para elaborar los productos para su distribución a las escuelas con el fin de apoyar a los profesores en el desarrollo de su práctica de evaluación formativa. Leahy y Wiliam desarrollaron dos paquetes de materiales, incluyendo clips de video y agenda, folletos y notas para el líder del grupo.¹¹⁵ Sin embargo, los investigadores reconocen que la existencia de estos materiales para el desarrollo profesional no garantiza su uso, lo cual depende de que las escuelas estén dispuestas a dar prioridad a la aplicación de la evaluación formativa.

Transformando las prácticas de evaluación de los profesores en Escocia

En Escocia, el proyecto de evaluación formativa fue parte de un proyecto más grande, el programa "Evaluación es para el Aprendizaje" (AifL, "*Assessment is for Learning*"), que considera el sistema de evaluación completo - los registros de los alumnos, la planificación personal, el sistema de monitoreo y la evaluación escolar, así como la evaluación formativa y la evaluación sumativa a nivel de sala de clases. El proyecto de evaluación formativa utilizó una variedad de actividades de desarrollo profesional, incluyendo talleres locales y conferencias nacionales en que las presentaciones fueron

114 Leahy, S. and Wiliam, D. (2012) op cit, p 54-5

115 Leahy, S. and Wiliam, D. (2009) *Embedding Assessment for Learning – A Professional Development Pack*. London: Specialist Schools and Academies Trust.

Leahy, S. and Wiliam, D. (2010) *Embedding Assessment for Learning – Pack 2*. London: Specialist Schools and Academies Trust.

hechas por los profesores que ya estaban involucrados en la implementación de la evaluación formativa en Inglaterra y más tarde por los profesores escoceses en la medida que progresaba el proyecto. Hubo también discusiones menos formales, pero sin embargo planificadas, discusiones entre los profesores locales y con los desarrolladores de la Autoridad Educacional ("*Educational Authority*"). También se hizo alguna referencia a las publicaciones del Grupo de Reforma de la Evaluación (*Assessment Reform Group*).¹¹⁶ Aunque los investigadores visitantes fueron incluidos en el programa de desarrollo profesional hay una clara impresión de que su participación fue decidida, diseñada y controlada en Escocia, en lugar de ser una colaboración formal con los que desarrollan la evaluación formativa en Inglaterra o en otro lugar.

Diseñadas o no, con las diversas necesidades de los diferentes profesores en mente, el conjunto de actividades provistas por el proyecto, sirvieron bien a las diversas formas en que los profesores implementaron la evaluación formativa en su propia práctica. Se identificaron tres enfoques principales

- *Ensayo y ajuste*: donde los profesores empezaron probando estrategias sugeridas en las publicaciones o por los profesores que ya estaban utilizando la evaluación formativa. Las estrategias se ajustaron a la luz de la experiencia durante los ensayos.
- *Revisión crítica y discusión previa al ensayo*: ideas para la implementación fueron discutidas con los colegas para aumentar la comprensión antes del ensayo.
- *Comenzando desde los objetivos y las ideas*: los profesores desarrollaron formas de incorporar en su práctica las ideas de la evaluación formativa proveniente de su experiencia compartida, más que el como lo había hecho el primer grupo involucrado en el KMOFAP.

Aunque hubo una impresión general de una implementación exitosa del proyecto por parte de los involucrados, el informe¹¹⁷ comenta sobre la aparente renuencia de algunos profesores a interesarse en las teorías de aprendizaje para entender por qué las estrategias funcionaban en el mejoramiento de los aprendizajes. También se observó que muy pocos profesores desarrollaron su práctica hasta el punto de permitir a sus estudiantes tomar parte en las decisiones acerca de las metas de aprendizaje. Si bien los estudiantes estaban tomando más iniciativa en la solución de problemas y se informó que estaban más involucrados con su pensamiento y más claros acerca de lo que deberían estar aprendiendo, los profesores mantenían un control sobre los objetivos de la lección y las metas a lograr. Es posible que estas dos cuestiones estuviesen relacionadas y que los profesores permitirán a los estudiantes asumir una mayor responsabilidad en su aprendizaje cuando ellos mismos desarrollen mayor comprensión de por qué la evaluación formativa funciona en términos de cómo aprenden los estudiantes. Como mostró el proyecto KMOFAP, esto lleva tiempo.

La importancia de la indagación en el desarrollo profesional de la evaluación formativa

Tomando en consideración una serie de iniciativas de investigación y desarrollo destinadas a ayudar al profesor o profesora a desarrollar estrategias de evaluación formativa en varios países, Pedder y James¹¹⁸ han identificado el aprendizaje profesional colaborativo como el factor más estrechamente relacionado con el uso de algunas estrategias de evaluación formativa. Este aprendizaje colaborativo puede adoptar diferentes formas, desde conversaciones informales entre los profesores, o profesores visitando sus respectivas salas de clases y discutiendo sus observaciones, hasta el más formal "Estudio de Lección" desarrollado en Japón. También concluyeron que debe colocarse un mayor

116 See www.assessment-reform-group.org/publications

117 Hayward, L. & Spencer, E. (2010) The Complexities of Change: formative assessment in Scotland, *The Curriculum Journal*, 21 (2) 161-177. Routledge

118 Pedder, D. and James, M. (2012) Professional learning as a condition for assessment for learning. In (ed) J. Gardner *Assessment and Learning*. 2nd edn. London: Sage, pp 33-48.

énfasis en las oportunidades ofrecidas a los docentes para utilizar los resultados de investigación relevantes y llevar a cabo la investigación durante su práctica:

En la medida que los docentes están preparados y comprometidos a participar en el arriesgado asunto de problematizar su propia práctica, buscando evidencia para evaluar con el fin de juzgar donde se requiere hacer cambios, y luego a actuar según sus decisiones, entonces ellos están incorporando la evaluación para el aprendizaje (evaluación formativa) a su propio aprendizaje profesional.¹¹⁹

Desarrollando la evaluación sumativa por los profesores de educación media

En un intento por encontrar un enfoque hacia la evaluación sumativa que tenga menos efectos negativos sobre el docente y los estudiantes que las pruebas nacionales que se aplican en Inglaterra, los investigadores y asesores de dos autoridades locales trabajaron con los profesores para desarrollar el uso del juicio docente para la evaluación sumativa.¹²⁰ El objetivo era encontrar métodos y procesos para garantizar la comparabilidad de los juicios entre los docentes y las escuelas. En su fase piloto, el proyecto trabajó con un grupo pequeño de profesores de inglés y de matemáticas de estudiantes de 8° Año (edad 13 años). Los profesores seleccionados conocían muy bien las prácticas de evaluación formativa. El proyecto se centró en el proceso de hacer juicios sumativos y en cómo la evidencia pueden convertirse en juicios. El desarrollo fue iniciado por los profesores identificando "qué significa ser bueno en esta asignatura para un estudiante de 8° Año?" En la segunda fase, durante el año escolar 2005-2006, las nuevas prácticas se pusieron a prueba y se adaptaron. La tercera fase incluyó que los profesores difundieran las ideas y prácticas a través de sus departamentos.

Los resultados del proyecto piloto, derivados de notas de campo, observaciones de clase, entrevistas y registros de reuniones, indican cierta confusión entre evaluación formativa y sumativa y la aceptación más que el desafío de la calidad de las pruebas actuales. También se encontraron diferencias en las reacciones de los profesores de matemáticas e inglés. Según se reportó, los profesores de inglés apreciaban el uso del portafolio, colocaban mayor énfasis en hablar y escuchar, y estaban introduciendo una pieza de trabajo "controlada" en la cual los estudiantes trabajaban solos. Los profesores de matemáticas prefirieron mantener la utilización de sus pruebas (algunos pusieron en duda la necesidad de un cambio) para mejorarlas, o bien introducir "tareas alternativas de evaluación" en lugar del enfoque más holístico, que fue el preferido por los profesores de inglés. Sin embargo, se encontró que el enfoque de las matemáticas no ofrecía oportunidades a los alumnos para mostrar un rango más amplio de logros, lo que sorprendió a los profesores cuando se dieron cuenta de ello.

Desarrollando la evaluación sumativa por los profesores de educación básica

Desde 1995, se ha esperado que los profesores en Inglaterra evalúen a sus estudiantes en el último año (6° nivel, 11 años de edad) de la escuela primaria, al mismo tiempo que se administran las pruebas nacionales en inglés, matemáticas y ciencias. En teoría, la evaluación por parte de los profesores estaba destinada a proporcionar datos acerca de un rango más amplio de aspectos de las asignaturas que aquellos que podrían ser incluidos en las pruebas. En la práctica, muchos profesores basaron sus evaluaciones en dar a los estudiantes test nacionales pasados, anulando así el propósito de informar sus juicios y perdiendo la oportunidad de usar los datos recogidos en el curso del trabajo regular para evaluar tanto formativa como sumativamente.

119 *ibid* p 41/2

120 Black, P., Harrison, C., Hodgen, J., Marshall, B. and Serret, N (2006) Riding the interface: an exploration of the issues that beset teachers as they strive for assessment systems. BERA conference paper see <http://www.kcl.ac.uk/sspp/departments/education/research/crestem/assessment/riding.pdf>

En 2006, se inició un pequeño proyecto¹²¹ en un intento de encontrar un modelo alternativo para reglamentar la evaluación y la moderación hecha por los docentes al final de la educación primaria. Con este fin se realizaron reuniones con un pequeño grupo de profesores del nivel 6° Año, durante el período en que estaban realizando el proceso de emitir los juicios evaluativos de fin de año. Estos juicios debían ser reportados decidiendo el "nivel" alcanzado por cada estudiante. Los "niveles" se identifican con las Metas de Logro del Currículo Nacional para cada asignatura (véase el ejemplo en el Recuadro 19, página 75.) Las reuniones permitieron a los profesores reflexionar sobre su experiencia y las diferentes formas de decidir sobre los niveles y moderar los procesos. Se puso de manifiesto en las primeras reuniones que los profesores estaban usando muy poca evidencia para hacer sus juicios, posiblemente como resultado de imitar las pruebas al decidir qué tomar en cuenta. Emitir juicios de niveles fue visto como problemático ya que las descripciones de los niveles abarcaban dos años de trabajo. Había un deseo de crear sub-niveles, pero a la vez reconocimiento de que era difícil establecer una evidencia firme para los pasos intermedios entre los niveles. En vez de eso, los docentes trabajando con los asesores produjeron algunas propuestas para ampliar la variedad de evidencia utilizada y algunos principios para guiar la práctica de emitir juicios acerca de los niveles alcanzados.

Estos materiales fueron puestos a disposición a un grupo mayor de profesores del nivel de 6° Año de 24 escuelas que participaron en el proyecto en 2007. Los profesores obtuvieron evidencia en un período de dos semestres y llevaron ejemplos a dos reuniones en las que sus juicios fueron moderados a través de la discusión con otros profesores. Las reuniones no sólo les dieron experiencia directa de moderación, sino también la oportunidad de compartir sus experiencias y discutir la selección de la evidencia. Ellos ampliaron la base de evidencia de sus juicios y comenzaron a ver que los alumnos podrían tener un rol en su propia evaluación. La reacción de los profesores a la experiencia de la moderación colaborativa fue altamente positiva.¹²² Ellos concluyeron que todos los profesores deberían tener esas oportunidades. Mientras que aquellos implicados valoraron para sí mismos la experiencia de desarrollar materiales orientadores, sugirieron que para una mayor diseminación el proceso debería ser "formalizado", lo que inevitablemente haría que fuera más "de arriba hacia abajo" para otros.

Cambiando la pedagogía y las prácticas de evaluación en Francia

Los cambios considerables en la educación francesa introducidos entre 2006 y 2009, que se mencionan en el Capítulo 6 (página 78), requirieron que a las escuelas implementaran una base común de habilidades, así como de conocimiento. La especificación de las habilidades fue una novedad, y en ciencias requirió que algunos profesores tuviesen que cambiar su aproximación a la enseñanza al mismo tiempo que sus prácticas de evaluación. El proyecto *La main à la pâte* ha proporcionado ayuda a las escuelas para estos dos cambios, de acuerdo a las prácticas actuales en las escuelas. En un distrito a un grupo de profesores de educación básica se les ayudó a desarrollar módulos de enseñanza sobre un tema en particular haciendo uso progresivo de las habilidades de indagación para los niveles de preescolar, 1° y 2° grado, y los niveles de 3° a 5° grado.¹²³ Cada módulo contiene una sección sobre evaluación. Las pruebas fueron también construidas y aplicadas a una muestra de estudiantes al final de la escuela básica (5° grado) con el fin de evaluar sus habilidades. El desempeño no fue reportado como notas o calificaciones, sino como "trabajando a través de la indagación", "avanzando hacia la indagación" y "todavía no trabajando a través de la indagación". Informes no publicados muestran que a los tres años el número de alumnos que no trabajan a través de la indagación disminuyó del 29% al 8%.

121 Birmingham City Council Advisory and Support Service (2005) *Effective Assessment*

122 Harlen, W. (2010) Professional learning to support teacher assessment, in Gardner, J. et al *Developing Teacher Assessment*. Maidenhead, England: Open University Press page 100-129.

123 (see http://www.crdp-montpellier.fr/cd66/map66/projets_federatifs/air/index.php).

En otro distrito, donde ECBI ya se estaba aplicando, las escuelas tuvieron acceso a tareas como las descritas en el Ejemplo 7 (página 79) para que pudiesen evaluar el progreso de sus estudiantes en habilidades de indagación y conocimientos. Aunque las pruebas no fueron diseñadas como evaluación sumativa, los profesores utilizaron los resultados, así como las observaciones de las acciones de los estudiantes y otras evidencias con el fin de completar la libreta de registro de cada estudiante al final del año. El efecto de facilitar el acceso a las tareas y la guía para el uso de las evidencias de los estudiantes de los cuadernos de ciencias, presentaciones y acciones (ver página 78) se demoró en aparecer. No se observó mucho cambio en el primer año, pero a partir de entonces, el uso de las tareas de evaluación y orientación se extendió rápidamente a una gran proporción (un 75%) de las escuelas del distrito.

Desarrollando prácticas de evaluación a través de materiales curriculares en Australia¹²⁴

En Australia, el proyecto Conexiones Primarias ("*Primary Connections*") tiene como objetivo desarrollar los conocimientos de los estudiantes, las habilidades, la comprensión y capacidades tanto en ciencias como en lectura y escritura. Promueve el enfoque indagatorio a través de lecciones estructuradas en cinco fases, 5E: Involucrarse (*Engage*), Explorar, Explicar, Elaborar y Evaluar. El desarrollo del proyecto ha sido financiado por el Gobierno Australiano desde el 2004 al 2012, y lo usa, de una manera u otra, el 56% de las escuelas primarias australianas. El programa tiene dos componentes principales: un programa de aprendizaje profesional y un conjunto de 31 unidades curriculares, que cubren el currículum australiano de ciencias desde la educación preescolar hasta el 6° Año. Las unidades curriculares modelan las prácticas indagatorias y la evaluación para y del aprendizaje. El programa de aprendizaje profesional incluye un taller sobre evaluación y varios DVDs de entrenamiento que han sido producidos para proporcionar ejemplos de la práctica en las aulas.

Se sugieren estrategias que puedan proporcionar información tanto para la evaluación formativa y para la evaluación sumativa. Un conjunto de preguntas del profesor se incluyen en las distintas fases de la lección para provocar el pensamiento de los estudiantes y hacer que sus ideas sean accesibles a los profesores y a los estudiantes, de manera que el aprendizaje pueda ser supervisado. También se incluyen estrategias para la evaluación de pares y de autoevaluación. El apéndice de cada unidad resume las oportunidades de evaluación para cada fase y las rúbricas de evaluación están disponibles en el sitio web para ayudar a los profesores a monitorear la comprensión conceptual de los estudiantes y las habilidades de indagación.

Un informe de retroalimentación profesores que participaron en el proceso de prueba¹²⁵ señala que los propósitos de la fase *Evaluar* son los siguientes:

- Proporcionar una oportunidad a los estudiantes de revisar y reflexionar sobre su aprendizaje y las nuevas comprensiones y habilidades
- proporcionar evidencia para los cambios en la comprensión, las creencias y las habilidades de los estudiantes.

En relación con el primero hubo una sólida evidencia de que los estudiantes revisaron su comprensión conceptual, pero no sus habilidades, aunque otras evidencias mostraron que ellos utilizaron una variedad de habilidades. Las respuestas de los profesores en relación con el segundo propósito eran escasas y sugerían que este aspecto de la fase no había sido implementado por muchos profesores. Los que lo hicieron, emplearon una serie de estrategias que aparecían servir a los dos propósitos de la fase *Evaluar*, pero nuevamente sólo en relación con la comprensión conceptual. Las estrategias

¹²⁴ Primary Connections <http://science.org.au/primaryconnections/>

¹²⁵ Skamp, K. (2012) Trial-teacher feedback on the implementation of Primary Connections and the 5E model. Australian Academy of Science <http://www.science.org.au/primaryconnections/research-and-evaluation/teaching-ps.html>

incluyeron pequeños cuestionarios, escritura de un artículo de prensa, la creación de un animal imaginario basado en criterios, dibujos, diagramas, juegos de palabras, mapas conceptuales, caricaturas conceptuales, juegos de roles y presentaciones. Los profesores señalaron que una serie de tareas, incluyendo cuestionamientos individuales y situaciones nuevas, eran necesarias para evaluar la comprensión de los estudiantes, lo que generaba demandas de altas exigencias a los tiempos de los profesores.

En general, si bien la evaluación por parte de los profesores no ha sido fuerte, la evidencia de la investigación muestra que algunos profesores están gradualmente avanzando hacia mejores prácticas de evaluación y están construyendo su confianza. Los procedimientos que han resultado ser efectivos en mejorar las prácticas de evaluación de los docentes incluyen:

- recursos curriculares que integran ejemplos de evaluación fundados en la investigación
- aprendizaje profesional que proporciona la teoría de por qué el enfoque funciona
- DVDs didácticos que muestran cómo se ve en las aulas reales
- experiencias de enseñanza positivas donde el gozo de los estudiantes y la evidencia del aprendizaje conducen al gozo de los profesores y a la motivación para participar en el cambio en la práctica docente.

Implicancias para la evaluación de ECBI

Hay algunas lecciones que aprender de estos intentos de cambiar la evaluación que se pueden aplicar para el aprendizaje y la enseñanza en cualquier asignatura. Antes de considerar lo que puede ser necesario en el caso de la evaluación de la educación en ciencias basada en la indagación, es útil recordar que, dado que no toda la enseñanza de la ciencia implicará indagación, debe haber una variedad de tareas y procedimientos de evaluación que se correspondan con la variedad de diferentes objetivos de aprendizaje. Como se ha mencionado antes, hay un lugar para la instrucción directa en las convenciones, el vocabulario y los conocimientos básicos que pueden ser controladas por pruebas o pequeñas interrogaciones preparadas por los profesores. Pero no toda evaluación debe ser del tipo en que la que las respuestas pueden ser recordadas. De lo contrario, no habrá ninguna información sobre las habilidades indagatorias de los estudiantes y la comprensión científica. Nuestro foco aquí se centra en los elementos esenciales de la evaluación que tienen que ver con los objetivos de ECBI.

Lamentablemente, el uso de procedimientos de evaluación que se ajustan a los propósitos de las metas en ECBI no está muy diseminado y casi no hay experiencia de lo que se necesita para cambiar las prácticas de evaluación para ajustarse mejor a estos propósitos. Proporcionar datos sobre la serie de competencias que constituyen las metas del ECBI, requiere de una serie correspondiente de procedimientos de evaluación - algunos que implican observar las acciones de los estudiantes, otros permitir que los estudiantes expresen sus ideas en discusiones, así como en la escritura y todos ofreciendo a los estudiantes la oportunidad de usar habilidades de indagación científica y razonamiento. Como se señaló en el Capítulo 6, los ejemplos de tales procedimientos existen, pero no son ampliamente utilizados en la educación científica. Sin embargo, por extrapolación de los procedimientos que han sido exitosos en otros dominios, es posible encontrar algunas pistas sobre qué ayuda a los profesores a cambiar su práctica de evaluación en ciencias.

Hay cuatro elementos claves en lograr un cambio en las prácticas de evaluación a nivel de clase y de la escuela que se aplican tanto a la evaluación formativa como en la sumativa:

- Motivación para cambiar
- Metas a alcanzar
- Oportunidades para los profesores de discutir, comparar y compartir soluciones
- Medios para evaluar el cambio en la práctica.

La *motivación* para realizar el cambio viene del descontento con la práctica actual y del reconocimiento que puede haber una mejoría. En la experiencia del proyecto KMOFAP (página 86), un factor clave para persuadir a los profesores a realizar el esfuerzo de introducir la evaluación formativa fue la evidencia de la investigación sobre el impacto en los logros de los estudiantes. Un factor para mantenerla en funcionamiento una vez iniciada, fue la evidencia del impacto que los cambios en su enseñanza generaron en sus propios estudiantes. En el caso de la evaluación sumativa, la motivación al cambio puede provenir de suministrar a los profesores evidencia que muestre que la evaluación sumativa convencional (la que se realiza a través de las pruebas convencionales) no les está dando a sus estudiantes la oportunidad de demostrar lo que son capaces de hacer y comprender. El involucrar a los docentes con más investigación como aquella realizada por Dolin y Krogh, que fue mencionada en el Capítulo 6 (página 70), podría influir para motivar el cambio. La evidencia de que el procedimiento utilizado en la evaluación puede hacer una diferencia significativa, es un primer paso para que los profesores se den cuenta de que lo que hace la diferencia es la visión del aprendizaje que subyace a los métodos de evaluación. Un enfoque que conduce a una evaluación más consistente con el entorno socio-cultural que sustenta ECBI, permite a los estudiantes demostrar sus capacidades en un grado mucho mayor que pruebas individuales o tareas.

Metas a alcanzar significa que los docentes y las escuelas tienen una idea clara de qué tipo de cambio se necesitará, pero tienen que desarrollar por sí mismos los detallados pasos requeridos para provocar ese cambio que debe ser apropiado a sus circunstancias particulares. En el caso de la evaluación formativa, se les puede dar información a los profesores acerca del rango de estrategias eficaces, pero adoptar diferentes formas de implementación en una sala de clases en particular. Para cambiar la evaluación sumativa, los profesores pueden ser ayudados a refinar sus ideas, sobre cómo identificar el logro y el progreso y cómo asegurar que los estudiantes tengan la oportunidad de mostrar sus logros y progreso.

Las oportunidades para discutir, comparar y compartir soluciones aseguran que, aunque adaptadas a las situaciones particulares, las prácticas mantengan un foco en las metas comunes. La discusión con otros que intentan conseguir los mismos efectos, pero en diferentes contextos, provee a los profesores con ideas que no se les hubiesen ocurrido trabajando solos. La indagación colaborativa funciona para ayudar a los profesores a construir y reconstruir ideas, tal como lo hace en el caso de los estudiantes. Es importante que esta colaboración y reflexión compartida esté disponible para todos. Algunos de los proyectos que se iniciaron con el desarrollo profesional colaborativo de los profesores alrededor de las nuevas prácticas de evaluación condujeron a la producción de materiales que los profesores consideraron podrían ser necesitados por otros. En tales casos, lo que comienza como transformación por unos pocos, termina como la transmisión para muchos. Evitar esto requiere el "delicado acto de equilibrio" identificado por Leahy y Wiliam, para evitar el peligro de los modelos de cambio basados en la transmisión - la suposición de que la práctica se puede cambiar siguiendo una receta proporcionada por los demás, incluso si esos otros son profesores.

Los *medios para evaluar los cambios* en la práctica de la evaluación, permiten a los profesores continuar con el desarrollo y ajustar sus prácticas a las nuevas circunstancias. Lo que no debe cambiar (o sólo en el largo plazo) son los principios que guían y proporcionan la justificación de las decisiones. A menos que se entiendan las razones del cambio, las técnicas serán seguidas en forma ciega y no serán adaptadas - como se requiere - para las diferentes situaciones, y con el tiempo eventualmente resultarán menos útiles. Es posible que tal como en el caso de los profesores del proyecto KMOFAP (página 86), el interés de saber *por qué* las nuevas prácticas funcionan se desarrolle después de ver que efectivamente funcionaron. Sin embargo, ya sea que la razón subyacente se presente más temprano o más tarde en el proceso de hacer los cambios, en algún momento los profesores necesitan tener una razón para adoptar las prácticas nuevas y abandonar las antiguas.

La lista del Recuadro 21 surgió como un ejemplo de principios relacionados a la práctica de la evaluación, en el marco de un proyecto que estudió los procesos involucrados al momento de hacer cambios en la práctica de la evaluación, especialmente en cuanto a colocar el rol de los docentes en el centro de evaluación¹²⁶.

Recuadro 21: Principios de las prácticas de evaluación del aprendizaje

- 1 La evaluación de cualquier tipo debería en último término mejorar el aprendizaje.
- 2 Los métodos de evaluación deberían permitir el progreso en todos los objetivos de aprendizaje importantes a ser facilitados y reportados.
- 3 Los procedimientos de evaluación deberían incluir procesos explícitos para asegurar que la información es válida y tan confiable como necesaria para su propósito.
- 4 La evaluación debería promover una comprensión pública de los objetivos de aprendizaje relevantes para la vida actual y futura de los estudiantes.
- 5 La evaluación de los objetivos de aprendizajes debería ser tratada como aproximaciones, susceptibles de errores inevitables.
- 6 La evaluación debería ser parte de un proceso de enseñanza que permita a los estudiantes comprender los propósitos de su aprendizaje y cómo será juzgada la calidad de sus logros.
- 7 Los métodos de evaluación deberían promover un compromiso activo de los estudiantes en su aprendizaje y su evaluación.
- 8 La evaluación debería permitir y motivar a los estudiantes a mostrar lo que son capaces de hacer.
- 9 La evaluación debería combinar información de diferente tipo, incluyendo la autoevaluación de los estudiantes, para informar las decisiones acerca de los aprendizajes y los logros de los estudiantes.
- 10 Los métodos de evaluación debería cumplir con estándares que reflejen un amplio consenso sobre calidad en todos los niveles, desde la práctica en la sala de clases hasta la política nacional.¹²⁶

El último de estos principios se refiere a los estándares de calidad que se pueden utilizar para evaluar la práctica. El proyecto que identificó los principios también propuso algunos estándares que se deben cumplir por parte de quienes toman las decisiones sobre la evaluación en la sala de clases, a nivel de la escuela, a nivel de la autoridad local y como parte de la política educativa nacional. Para el estudiante individual, lo que sucede en la sala de clase tiene el mayor impacto en el aprendizaje; por ello hay estándares que deben cumplirse si es que la evaluación es para ayudar al aprendizaje. Pero, lo que sucede en sala de clases depende de lo que pasa en la escuela en términos de la política de evaluación y el discurso general sobre cómo la evaluación debe ayudar o informar o medir los logros de los estudiantes. A su vez la política y la práctica de la escuela están influenciadas por la orientación

¹²⁶ Gardner, J., Harlen, W., Hayward, L. and Stobart, G. with Montgomery, M. (2010). *Developing Teacher Assessment*. Maidenhead: Open University Press pp 48-51.

de la autoridad local y por los requerimientos de la política nacional. Por lo tanto, el proyecto identificó estándares que deben cumplirse en la práctica dentro de cada una de estas comunidades.

Los estándares se reproducen en las Tablas 1 - 4. En cada caso hay estándares para la evaluación del aprendizaje en general, para la evaluación formativa y la evaluación sumativa.

Tabla 1: Estándares para la Práctica de Evaluación del Aprendizaje en el Aula

Evaluación del Aprendizaje en General	Uso Formativo de la Evaluación	Uso Sumativo de la Evaluación
1 La evaluación utiliza una serie de métodos que permiten abordar los diversos objetivos de aprendizaje y su progresión. 2 Los métodos utilizados abordan las habilidades, el conocimiento o comprensión que se está evaluando, sin restringir el alcance del currículum. 3 La enseñanza proporciona a los estudiantes la oportunidad de demostrar lo que pueden hacer a través de tareas que abordan toda la serie de objetivos de aprendizaje. 4 Los profesores utilizan evidencia que proviene de su evaluación continua para: • ayudar el aprendizaje de los estudiantes; • resumir el aprendizaje en términos de comunicar criterios; • reflexionar y mejorar su enseñanza. 5 Los profesores desarrollan su práctica de evaluación a través de una variedad de actividades de aprendizaje profesional incluyendo la reflexión y el compartir experiencias con colegas	1 Los profesores reúnen evidencia del aprendizaje de sus alumnos a través de preguntas, observación, discusión y estudio de los productos relevantes para los objetivos de aprendizaje. 2 Los profesores involucran a los estudiantes en la discusión de los objetivos de aprendizaje y los estándares que se esperan de su trabajo. 3 Los profesores utilizan la evaluación para avanzar el aprendizaje de los estudiantes: • adaptando el ritmo, el desafío y el contenido de las actividades • retroalimentando a los estudiantes sobre cómo mejorar • proporcionando tiempo a los estudiantes para reflexionar sobre y evaluar su propio trabajo. 4 Los estudiantes utilizan la evaluación para avanzar su aprendizaje a través de: • conocimiento y utilización de los criterios de los estándares del trabajo que esperan lograr • dar y recibir comentarios de sus compañeros sobre la calidad de su trabajo y cómo mejorarlo • reflexionar sobre cómo mejorar su trabajo y asumir su responsabilidad por éste	1 Los profesores basan sus juicios de los logros de aprendizaje de los estudiantes en una variedad de tipos de actividades adecuadas a la asignatura y a la edad de los alumnos, las que pueden incluir pruebas o tareas específicas de evaluación. 2 La evaluación de los resultados de aprendizaje se basa en una gran variedad de tareas que permite a los estudiantes demostrar lo que significa ser "bueno" con respecto a un trabajo en particular. 3 Los profesores discuten con otros, sobre el trabajo de los estudiantes con el fin de alinear los juicios de los niveles o grados cuando éstos se requieren. 4 Los estudiantes están conscientes de los criterios por los que su trabajo es juzgado durante un período de tiempo. 5 Los estudiantes están conscientes de la evidencia utilizada y cómo los juicios de los logros de aprendizaje se han realizado. 6 Los alumnos reciben ayuda para utilizar los resultados de la evaluación para mejorar su aprendizaje.

Tabla 2: Estándares para ser usados por el Equipo de Gestión de la Escuela

Evaluación del Aprendizaje en General	Uso Formativo de la Evaluación	Uso Sumativo de la Evaluación
1 Existe una política de la escuela para la evaluación que refleja los estándares anteriores para la práctica en el aula. 2 La política es regularmente discutida y revisada para reflejar una práctica en desarrollo. 3 Los profesores tienen la oportunidad de mejorar sus prácticas de evaluación a través del aprendizaje profesional y la colaboración. 4 Se pone tiempo a disposición de los profesores para discutir, reflexionar y de vez en cuando para observar la práctica de evaluación de unos y otros. 5 La política y la práctica de la escuela en evaluación es comunicada a los padres y apoderados.	Los profesores colaboran en el desarrollo de sus prácticas cuando: • comunican los objetivos y criterios de calidad a sus alumnos • ayudan a sus alumnos a participar en la autoevaluación y evaluación de pares • proveen retroalimentación para ayudar el aprendizaje • permiten a los estudiantes asumir responsabilidad por su trabajo.	1 Los profesores pueden utilizar una variedad de métodos de evaluación libres de la presión por el uso de alto impacto de los resultados. 2 Los profesores participan en el desarrollo de procedimientos de aseguramiento de la calidad para maximizar la consistencia en sus juicios. 3 Los logros de los estudiantes se discuten en términos de lo que pueden hacer y no sólo en términos de niveles o grados. 4 Un sistema manejable para la mantención de registros está en funcionamiento para realizar un seguimiento e informar sobre el aprendizaje de los estudiantes. 5 Los padres y apoderados reciben informes escritos y orales que identifican los próximos pasos de sus hijos y proporcionan información sobre los procesos de evaluación para garantizar la confianza en la evaluación de los profesores. 6 Los juicios sumativos se requieren sólo cuando son necesarios para verificar y reportar el progreso.

Tabla 3: Estándares para ser usados en la Inspección Nacional y Dispositivos de Asesoría

Evaluación del Aprendizaje en General	Uso Formativo de la Evaluación	Uso Sumativo de la Evaluación
1 Las políticas y prácticas de las escuelas en materia de evaluación son revisadas en relación con los estándares anteriores. 2 Los procedimientos de inspección aseguran que las escuelas evalúen sus prácticas de evaluación del aprendizaje y desarrollen planes de acción para su mejoramiento. 3 Existen oportunidades para las escuelas de compartir y desarrollar prácticas de evaluación. 4 Se dispone de desarrollo profesional para desarrollar políticas y mejorar las prácticas de evaluación. 5 Se dispone de recursos para que las escuelas puedan tomar medidas hacia el mejoramiento de las prácticas de evaluación.	1 El uso de la evaluación por parte de la escuela, para apoyar el aprendizaje, se incluye como un factor clave en la evaluación de la efectividad de las escuelas. 2 Existe ayuda disponible para que las escuelas aseguren que todas las áreas de logro se beneficien del uso formativo de la evaluación. 3 Las escuelas son estimuladas a desarrollar el uso formativo de la evaluación.	1 Las escuelas son ayudadas a desarrollar planes de acción basados en la auto-evaluación a través de una serie de indicadores que van más allá de los niveles de logro de los estudiantes. 2 El consejo a las escuelas sobre las políticas y prácticas de evaluación tiene en cuenta lo que se sabe acerca de la fiabilidad y la validez de los diferentes métodos de evaluación. 3 Las escuelas son ayudadas a utilizar los resultados de la evaluación para identificar las áreas para mejorar las oportunidades de aprendizaje.

Tabla 4: Estándares para ser usados en la Formulación de Política Pública

Evaluación del Aprendizaje en General	Uso Formativo Use de la Evaluación	Uso Sumativo de la Evaluación
1 Las políticas requieren que las escuelas y los asesores locales muestren cómo todas las evaluaciones están siendo utilizadas para ayudar al aprendizaje de los alumnos. 2 La introducción de nuevas prácticas en materia de evaluación se acompaña de cambios en la formación de los profesores y criterios de evaluación necesarios para su sostenibilidad. 3 Las escuelas son responsables del uso de la evaluación formativa y sumativa para maximizar el logro de las metas. 4 Los estándares nacionales de desempeño de los estudiantes son reportados como una serie de datos cualitativos y cuantitativos obtenidos en estudios de muestras representativas.	1 La evaluación para apoyar el aprendizaje está en el centro de los programas de gobierno para elevar los niveles de logro. 2 La formación inicial de profesores y los cursos de desarrollo profesional aseguran que los profesores tengan las habilidades para utilizar la evaluación para apoyar el aprendizaje. 3 El marco legal de inspección de las escuelas releva el uso de la evaluación para apoyar el aprendizaje. 4 Se estimula a las escuelas a evaluar y desarrollar el uso formativo de la evaluación.	1 La evaluación moderada por los profesores se usa para informar el desempeño de los estudiantes a lo largo de los años escolares obligatorios. 2 La moderación de los juicios de los docentes se requiere para asegurar una interpretación común de criterios al interior y a través de las escuelas. 3 El reglamento garantiza que las disposiciones para el uso de la evaluación sumativa es compatible con la práctica de utilizar la evaluación para ayudar al aprendizaje. 4 Los objetivos para la mejora de la escuela se basan en una serie de indicadores y se acuerdan a través de un proceso que combina la evaluación externa y autoevaluación interna.

Referencias Bibliográficas

- Alexander, R. (Ed) (2010) *Children, their World, their Education*. Final report and recommendations of the Cambridge Primary Review. London: Routledge.
- Alexander, R. (2004) *Towards Dialogic Teaching. Rethinking Classroom Talk*. Cambridge: Dialogos.
- Assessment Reform Group (ARG) (2002) *Assessment for Learning: 10 Principles*. www.assessment-reform-group.org
- Australian Government Department Education, Employment and Workplace Relations. *Science Education Assessment Resources (SEAR)* <http://cms.curriculum.edu.au/SEAR>
- Barnes, D. (1976) *From Communication to Curriculum*. Harmondsworth: Penguin.
- Birmingham City Council Advisory and Support Service (2005) *Effective Assessment*.
- Black, P. (1998) *Testing: Friend or Foe?* London: Falmer Press.
- Black, P., Harrison, C., Hodgen, J., Marshall, B. and Serret, N (2006) Riding the interface: an exploration of the issues that beset teachers as they strive for assessment systems. BERA conference paper . See <http://www.kcl.ac.uk/sspp/departments/education/research/crestem/assessment/riding.pdf>
- Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. (2003). *Assessment for Learning: Putting it into Practice*. Maidenhead England: Open University Press.
- Black, P. and Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21 (1). 5-13.
- Black, P. and Wiliam, D. (1998) Assessment and classroom learning, *Assessment in Education*, 5 (1) 7-74.
- Bransford, J.D., Brown, A. and Cocking, R.R. (eds) (2000) *How People Learn, Brain, Mind, Experience and School*. Washington, D.C.: National Academy Press.
- Bruner, J. (1996) *The Culture of Education*. Cambridge, MA: Harvard University Press.
- Budd-Rowe, M. (1974) Relation of wait-time and rewards to the development of language, logic and fate control: Part II, *Journal of Research in Science Teaching*, 11(4) 291-308.
- Butler, R. (1988) Enhancing and undermining intrinsic motivation: the effects of task-involving and ego-involving evaluation on interest and performance, *British Journal of Educational Psychology* 58, 1-14.
- Crossouard (2012) Absent presences: the recognition of social class and gender dimensions within peer assessment interactions, *British Educational Research Journal*, 38 (5) 731-748.
- Department for Education (2011). *The Framework for the National Curriculum. A report by the Expert Panel for the National Curriculum review*. London: Department for Education.
- DES, DENI and WO (1985) *APU Science in Schools Age 11 Report no 4*. London: HMSO.
- DES and WO (1988) *National Curriculum Task Group on Assessment and Testing: A Report*. London: HMSO.
- DES, DENI and WO (1981) *Science in Schools Age 11 Report no 1*. London: HMSO.
- Dewey, J. (1933) *How we think: A restatement of the relation of reflective thinking to the educative process*. Boston, MA: D.C. Heath.
- Dolin, J., & Krogh, L. B. (2010): The Relevance and Consequences of Pisa Science in a Danish Context. *International Journal of Science and Mathematics Education*, 8, 565-592.

- Dolin, J., Laursen, E., Raae, P. H., Senger, U. (2005). *Udviklingsprojekter som læringsrum. Potentialer og barrierer for skoleudvikling i det almene gymnasium. Gymnasiepædagogik nr. 54*, Syddansk Universitet. 232 s. (Development projects as learning arenas. Potentials and barriers for school development in the general upper secondary school. University of Southern Denmark). http://www.sdu.dk/Om_SDU/Institutter_centre/ikv/Formidling/Tidsskrifter/Gymnasiepædagogik/Udgaver Accessed Dec11, 2012.
- Ertl, H. (2006) Educational standards and the changing discourse on education: The reception and consequences of the PISA study in Germany. *Oxford Review of Education*, 32 (5) pp 619–634.
- Gardner, J., Harlen, W., Hayward, L. and Stobart, G. with Montgomery, M. (2010). *Developing Teacher Assessment*. Maidenhead: Open University Press.
- Gipps, C., McCallum, B. and Brown, M. (1996). Models of teacher assessment among primary school teachers in England, *The Curriculum Journal*, 7 (2) 167–183.
- Grigorenko, E. (1998) Mastering tools of the mind in school (trying out Vygotsky's ideas in classrooms), in (eds) R. Sternberg and W. Williams *Intelligence, Instruction and Assessment: Theory and Practice*. Mahwah, NJ: Erlbaum.
- Harlen, W. (2012a) The role of assessment in developing motivation for learning, in (ed) J. Gardner *Assessment and Learning*. London: Sage pp171–184.
- Harlen, W. (2012b) On the relationship between assessment for formative and summative purposes, in (ed) J. Gardner *Assessment and Learning*. London: Sage pp 87–102.
- Harlen, W. (2010) Professional learning to support teacher assessment, in Gardner, J. *et al Developing Teacher Assessment*. Maidenhead England: Open University Press page 100–129.
- Harlen (Ed) (2010) *Principles and Big Ideas of Science Education*. Available from www.ase.org.uk in English, from www.fondation-lamap.org in French, and from www.innovec.org.mx in Spanish.
- Harlen, W. (2007) *Assessment of Learning*. London: Sage.
- Harlen, W. (2006) *Teaching, Learning and Assessing Science 5 – 12*. 4th edn. London: Sage.
- Harlen, W. (2004) Trusting teachers' judgements: research evidence of the reliability and validity of teachers' assessment for summative purposes, *Research Papers in Education*, 20(3); 245–270.
- Harlen, W. and Qualter, A. (2009) *The teaching of Science in Primary Schools*. London: Routledge.
- Hattie, J. and Timperley, H. (2007) The power of feedback. *Review of Educational Research*, 77, 81–112.
- Hayward, L. (2010) Moving beyond the classroom, in J. Gardner et al *Developing Teacher Assessment*. Maidenhead, UK: Open University Press.
- Hayward, L. & Spencer, E. (2010) The Complexities of Change: formative assessment in Scotland, *The Curriculum Journal*, 21 (2) 161–177.
- IAP (2012) Taking Inquiry-Based Science Education into Secondary Education. Report of a global conference. <http://www.sazu.si/files/file-147.pdf>
- Jager, J.J., Merki, K.M., Oerke, B. and Holmeier, M. (2012) State-wide low-stakes tests and a teaching to the test effect? An analysis of teacher survey data from two German States, *Assessment in Education*, 19 (4) 451–467.
- James, M. (2012) Assessment in harmony with our understanding of learning: problems and possibilities, in (ed) J. Gardner *Assessment and Learning*, 2nd edn. London: Sage 187 – 205.
- Leahy, S. and Wiliam D. (2012) From teachers to schools: scaling up professional development for formative assessment, in (ed) J. Gardner *Assessment and Learning*, 2nd edn. London: Sage. 49–71.
- Leahy, S. and Wiliam, D. (2009) *Embedding Assessment for Learning – A Professional Development Pack*. London: Specialist Schools and Academies Trust.

- Leahy, S. and Wiliam, D. (2010) *Embedding Assessment for Learning – Pack 2*. London: Specialist Schools and Academies Trust.
- Linn, R. L. (2000) Assessments and accountability, *Educational Researcher*, 29 (2) 4-16.
- Masters, G. and Forster, M. (1996) *Progress Maps*. Camberwell, Victoria, Australia: ACER.
- Maxwell, G. (2004) 'Progressive assessment for learning and certification: some lessons from school-based assessment in Queensland.' Paper presents at the third conference of the Association of Commonwealth Examination and Assessment Boards, March Nidi, Fiji.
- Messick, S.(1989) Validity, in (ed) R. Linn *Educational Measurement* (3rd Edn) American Council on Education , Washington: Macmillan, pp 13-103.
- Michaels, S., Shouse, A.W. and Schweingruber, H.A (2008) *Ready, Set, Science! Putting research to work in K-8 Science Classrooms*, Washington: National Academies Press.
- Minner, D.D., Levy, A. J and , Century, J. (2010) Inquiry-Based Science Instruction—What Is It and Does It Matter? Results from a Research Synthesis Years 1984 to 2002, *Journal of Research in Science Teaching*, 47 (4) 474-496.
- National Research Council (2012) *A Framework for K-12 Science Education*. Washington DC: National Academies Press.
- Newton, P. E. (2012) Validity, purpose and the recycling of results from educational assessment, in (Ed) J. Gardner *Assessment and Learning* 2nd edition. London: Sage 264-276.
- Noble, T., Suarez, C., Rosebery, A., O'Connor, M.C. ,Warren, B. and Hudicourt-Barnes, J. (2012) "I never thought of it as freezing": How Students Answer Questions on large-scale science tests and what they know about science, *Journal of Research in Science Teaching*, 49 (6) 778–803.
- Nordenbo, S. E., Allerup, P., Andersen, H. L., Dolin, J., Korp, H., Larsen, M. S., et al. (2009). *Pædagogisk brug af test - Et systematisk review*. København: Aarhus Universitetsforlag. (In English: *Pedagogical use of tests – A systematic review*).
- Nuffield Primary Science Teachers' Guide *Materials*. (1995) London: Collins Educational.
- Nusche, D., Laveault, D., MacBeath, J. and Santiago, P. (2012) *OECD Reviews of Evaluation and Assessment in Education: New Zealand 2011*. Paris: OECD.
- OECD (2011) *Towards an OECD Skills Strategy*. Paris: OECD.
- OECD (2006) *PISA released items: Science*. Paris: OECD <http://www.oecd.org/pisa/38709385.pdf>
- OECD 2003, *The PISA 2003 Assessment Framework* Paris: OECD.
- OECD (2000) *Measuring Student Knowledge and Skills: A new Framework for Assessment*. Paris: OECD.
- Osmundson, E., Chung, G., Herl, H., Klein D. (1999) *Knowledge-mapping in the classroom: a tool for examining the development of students' conceptual understandings*. Los Angeles, California: National Centre for Research on Evaluation and Student Testing, University of California. www.cse.ucla.edu/Reports/TECH507.pdf
- Osborne, J., Simon, S. and Collins, S.(2003) Attitudes towards science: a review of the literature and its implications, *International Journal of Science Education*, 25, 1049-1079.
- Pedder, D. and James, M. (2012) Professional learning as a condition for assessment for learning. In (ed) J. Gardner *Assessment and Learning*. 2nd edn. London: Sage, pp 33-48.
- Pellegrino, J.W., Chudowsky, N. and Glaser, R. (Eds) (2001) *Knowing what Students Know The Science and Design and Educational Assessment*. Washington, DC: National Academy Press.
- Piaget, J (1929) *The Child's Conception of the World*. New York: Harcourt Brace.

- Pine, J., Aschbacher, P., Rother, E., Jones, M., McPhee, C., Martin, C., Phelps, S., Kyle, T. and Foley, B. (2006) Fifth graders' science inquiry abilities: a comparative study of students in hands-on and textbook curricula, *Journal of Research in Science Teaching* 43 (5): 467-484.
- Pollard, A and Triggs, P. (2000) *Policy, Practice and Pupil Experience*. London: Continuum.
- Pollard, A., Triggs, P., Broadfoot, P., Mcness, E. and Osborn, M. (2000) *What pupils say: changing policy and practice in primary education*. London: Continuum.
- Primary Connections <http://science.org.au/primaryconnections/>
- Pryor, J. and Lubisi, C. (2001) Reconceptualising educational assessment in South Africa –testing times for teachers, *International Journal for Educational Development*, 22 (6), 673-686.
- Roderick, M. and Engel, M. (2001) The grasshopper and the ant: motivational responses of low achieving pupils to high stakes testing. *Educational Evaluation and Policy Analysis* 23: 197-228.
- Sadler, D. R. (1989) Formative assessment and the design of instructional systems, *Instructional Science*, 18, 119-44.
- SEED (Scottish Executive Education Department) (2002) *How Good is Our School? Self evaluation using quality indicators*. Edinburgh: HMIE.
- Skamp, K. (2012) *Trial-teacher feedback on the implementation of Primary Connections and the 5E model*. Australian Academy of Science. <http://www.science.org.au/primaryconnections/research-and-evaluation/teaching-ps.html>
- Stobart, G. (2012) Validity in formative assessment, in (ed) J. Gardner *Assessment and Learning*. 2nd edn. London: Sage pp 233-242.
- Stobart, G. (2008) *Testing Times. The uses and abuses of assessment*. London: Routledge.
- Streeter, L., Bernstein, J., Foltz, P. and DeLand, D. (2011) *Pearson's Automated Scoring of Writing, Speaking, and Mathematics*. Pearson
<http://kt.pearsonassessments.com/download/PearsonAutomatedScoring-WritingSpeakingMath-051911.pdf>
- Tymms, P. (2004) Are standards rising in English primary schools? *British Educational Research Journal*, 30 (4) 477-94.
- Vygotsky, L.S. (1978) *Mind in Society: The Development of Higher Psychological Process*. Cambridge, MA: Harvard University Press.
- Watkins, C. (2003) *Learning: A Sense-Maker's Guide*. London: Association of Teachers and Lecturers.
- Welford, G., Harlen, W. and Schofield, B. (1985) *Practical Testing at ages 11, 13, and 15*. London: DES, WO and DENI.
- White, R. T. (1988) *Learning Science*. Oxford: Blackwell.
- Wiliam, D. (2001) Reliability, validity and all that jazz, *Education 3-13*, 29 (3): 17-21.
- Wiliam, D. (2009) An integrative summary of the research literature and implications for a new theory of formative assessment, in (eds) H. L. Andrade and G. J. Cizek, *Handbook of Formative Assessment*, New York: Taylor and Francis.
- Wiliam, D. Lee, C., Harrison, C. and Black, P. (2004) Teachers developing assessment for learning: impact on student achievement, *Assessment in Education*, 11 (1) 49-66.

