

Graham Gibbs
Claire Simpson

**CONDICIONES PARA UNA
EVALUACIÓN CONTINUADA
FAVORECEDORA DEL APRENDIZAJE**



OCTAEDRO



Cuadernos de docencia universitaria 13

Título original: *Conditions under which assessment support students' learning*

Traducción al castellano: Marta Fernández-Villanueva Jané

Primera edición: enero de 2009

© Graham Gibbs y Claire Simpson

© ICE y Ediciones OCTAEDRO, S.L.

Ediciones OCTAEDRO

Bailén, 5 - 08010 Barcelona

Tel.: 93 246 40 02 - Fax: 93 231 18 68

www.octaedro.com - octaedro@octaedro.com

Universitat de Barcelona

Institut de Ciències de l'Educació

Campus Mundet - 08035 Barcelona

Tel.: 93 403 51 75 - Fax: 93 402 10 61

La reproducción parcial o total de esta obra sólo es posible de manera gratuita e indicando la referencia de los titulares propietarios del copyright: ICE y Octaedro.

ISBN: 978-84-8063-934-7

Diseño y producción: Servicios Gráficos Octaedro



ÍNDICE

Resumen	5
Introducción	6
La influencia determinante de la evaluación	7
La importancia de las actividades y trabajos de curso	10
El declive de la evaluación formativa	13
La efectividad del <i>feedback</i>	15
Cómo influir a través de la evaluación sobre el volumen, la selección y calidad de lo estudiado	19
Condición 1. Ofrecer a los estudiantes suficientes actividades evaluables para ocupar el tiempo de estudio necesario y suficiente	19
Condición 2. Ocupar a los estudiantes con actividades evaluables que les orienten sobre cómo distribuir apropiadamente el tiempo y el esfuerzo en función de los aspectos más importantes de la asignatura	21
Condición 3. Afrontar una tarea a evaluar implica realizar unas actividades de aprendizaje productivas del tipo más apropiado	22
La influencia del <i>feedback</i> en el aprendizaje	25
Condición 4. Dar suficiente <i>feedback</i> , tanto con suficiente frecuencia como con el suficiente detalle	25
Condición 5. Centrar el <i>feedback</i> en la actuación de los estudiantes, en su aprendizaje y en las acciones que están bajo su control, en lugar de centrarlo en los propios estudiantes o en sus características	27
Condición 6. Ofrecer el <i>feedback</i> a tiempo para que sea recibido cuando todavía le importa a los estudiantes y puedan utilizarlo en un aprendizaje posterior o puedan recibir ayuda a tiempo	27
Condición 7. El <i>feedback</i> es apropiado en relación al propósito de la actividad y a los criterios de evaluación	29
Condición 8. El <i>feedback</i> es apropiado en relación a lo que entienden los estudiantes que tienen que hacer	31
Concepciones de la tarea que tienen los estudiantes	31
Concepciones del aprendizaje que tienen los estudiantes	31
Concepciones del conocimiento que tienen los estudiantes	32
La concepción del lenguaje de la disciplina por parte del estudiante	33
Condición 9. El <i>feedback</i> es recibido y tenido en cuenta	34

Condición 10. El estudiante actúa ante el <i>feedback</i>	35
Conclusión	37
Agradecimientos	38
Bibliografía.....	39

PRÓLOGO

Me comentaba un estudiante de una universidad politécnica española que había cursado un curso académico en una universidad norteamericana de prestigio, que su vida como estudiante cambió por completo en la universidad de acogida y volvió a cambiar cualitativamente cuando volvió a nuestro país. Entre los cambios que vivió allí destacaba los siguientes: más trabajo personal con menos asistencia presencial a clase; menos contenidos que aprender; más tiempo para vivir y disfrutar –más bienestar– como estudiante y un sistema de evaluación diferente –fluida y eficaz– que consistía en ofrecer información de manera continua sobre su proceso de aprendizaje. A todos nos apetece que no nos den más información de la estratégicamente necesaria para entender y comprender algo, que nos devuelvan información sobre cómo aprendemos con prontitud y claridad y disponer de tiempo para vivir y disfrutar además de trabajar y estudiar. Probablemente hubo otros factores que ayudaron a que su vida en aquella universidad fuera tan bien valorada un par de años después. Pero sea por la razón que fuere el estudiante seguía destacando que un sistema de evaluación que facilita continuamente información sobre cómo aprendes y una buena selección de contenidos en función de su valor estratégico y no de la masa que ocupan son auténticos factores de bienestar para el estudiante.

Creo que no sólo son factores de bienestar para el estudiante que quiere aprender sino que además son factores clave para la optimización de la calidad de la docencia en la universidad. Y no creo que sean consecuencia ni de la implantación del Espacio Europeo de Educación Superior, ni de ninguno de los documentos derivados de la declaración de Bolonia. Sencillamente son consecuencias lógicas de cualquier proyecto de mejora de la calidad de la docencia en la universidad. La mejora de la calidad de la docencia no se alcanza sólo estableciendo buenos objetivos y expresándonos en términos de resultados de aprendizaje y competencias. Requiere sobre todo ser riguroso al escoger los contenidos en función de su valor estratégico y cuidadoso en la evaluación. Conviene aprovechar la integración en el proceso de convergencia europea para repensar nuestra tarea docente y dedicar tiempo a estas cuestiones. Una vez ajustados y en los mejores casos construidos los planes

docentes de los nuevos grados y postgrados, conviene dedicar nuestro tiempo académico a escoger los mejores contenidos y practicar buenos sistemas de evaluación continuada.

Los cambios que acompañan al proceso de construcción del espacio Europeo de Educación Superior pueden aprovecharse para generar condiciones en los contextos de docencia universitaria que fomenten aprendizaje en profundidad. Entre estos cambios los hay que están generados por la dinámica del propio proceso de elaboración y propuesta de los nuevos títulos de grado y postgrado –si esta tarea no se ha limitado a un mero arreglo superficial en función de un ajuste de resultados en términos de recursos docentes por departamento– como por ejemplo la práctica de la planificación docente por materias y la programación por asignaturas. Otros vienen de la mano de la integración de las tecnologías de la información y la documentación al servicio del aprendizaje y la comunicación y permiten avanzar hacia sistemas de aprendizaje y docencia en la universidad de carácter híbrido que combinan situaciones presenciales con otras que no lo son. Otros están asociados a la propia noción de «buen nivel de aprendizaje» en una sociedad como la actual, que está caracterizada por un crecimiento exponencial de la información en los diferentes ámbitos del conocimiento que obliga a escoger de entre todos los contenidos disponibles y susceptibles de ser enseñados aquellos que generan mejor aprendizaje. Pero para lograr un mejor aprendizaje y que lo sea en profundidad es necesario que la planificación docente, la programación, la integración de tecnologías para el aprendizaje y la comunicación y la selección de contenidos estén acompañadas de sistemas de evaluación que fomenten aprendizaje fructífero a largo plazo y equipos docentes desarrollando tales tareas.

Se puede afirmar, porque existen evidencias que lo corroboran, que las asignaturas en las que los estudiantes obtienen mejores resultados son aquellas que se caracterizan por una estrecha colaboración entre el profesorado que las imparte. En cambio las asignaturas con peores calificaciones son aquellas en las que el profesorado se ocupa y concentra sólo en su asignatura. Lo que determina el resultado global del aprendizaje del estudiante no es la suma de espacios docentes estancos –por muy buena que sea la docencia en ellos– sino el trabajo conjunto del profesorado en el marco del plan docente de la titulación. Por ello el

trabajo en equipos docentes es clave. Los equipos docentes permiten la discusión y reflexión sobre la práctica del profesorado que lo conforma, una transferencia más ágil de la innovación que contribuye a la mejora de la calidad y facilita la evaluación conjunta.

Sin embargo el profesorado está habituado a desarrollar su tarea docente y evaluadora en clave individual y en el contexto de la asignatura. El cambio hacia una cultura docente que integre el trabajo del profesorado mediante equipos docentes no es fácil, pero seguro que es necesario. Por ello es conveniente generar espacios de formación y análisis en los que el profesorado que participa pueda reflexionar sobre su práctica docente y construir conocimiento sobre los procesos de aprendizaje, docencia y evaluación en la universidad de manera colaborativa. Espacios de formación e intercambio entre profesorado de disciplinas afines o de disciplinas diferentes que favorezcan capacidades y actitudes de cooperación entre los participantes y desarrollen sus competencias para escoger los mejores contenidos, comunicar mejor el conocimiento y conocer las funciones estratégicas de la evaluación que permitan a los docentes seleccionar las tácticas evaluativas más apropiadas.

Sobre esta última cuestión, la evaluación de los aprendizajes de los estudiantes, trata el texto que presentamos. Desde el Instituto de Ciencias de la Educación y mediante nuestra colección Cuadernos de docencia universitaria, nos proponemos con este cuaderno ofrecer un conjunto de reflexiones y una revisión sobre cuáles son las mejores condiciones para una evaluación continuada favorecedora del aprendizaje de los estudiantes.

El trabajo de Graham Gibbs y Claire Simpson que ahora presentamos es una excelente revisión de los estudios sobre la evaluación de los aprendizajes en la universidad. El texto propone un conjunto de condiciones bajo las cuales la evaluación favorece el aprendizaje y las argumenta con referencias teóricas, evidencias empíricas y experiencias prácticas. Marta Fernández Villanueva, profesora de la Facultad de Filología de la Universidad de Barcelona y hasta hace unos meses coordinadora del programa de asesoramiento en formación para la innovación docente del ICE de la UB ha traducido el texto y ha adaptado a nuestro lenguaje académico y contexto universitario las consideraciones y reflexiones

desarrolladas en el mismo. A Graham y Claire nuestro más sincero agradecimiento por su estudio y disponibilidad y a Marta por su buen trabajo.

Es un texto ágil que, como muy bien señalan sus autores, no trata en absoluto de sistemas de medición, trata del aprendizaje. Un texto que defiende y argumenta que lo importante en la evaluación es que favorezca un aprendizaje fructífero y que de la fiabilidad hay que ocuparse después. En el texto se afirma la influencia determinante de la evaluación en el aprendizaje de los estudiantes, se considera el declive de la evaluación formativa ante la limitación de recursos en algunas universidades, y se propone y argumenta diez condiciones para una evaluación continuada del aprendizaje.

Creo que es un excelente texto para fomentar el análisis, la reflexión y el debate entre el profesorado sobre el valor estratégico de la evaluación, las concepciones que los estudiantes tienen sobre el aprendizaje y la evaluación y para avanzar hacia tácticas y prácticas evaluativas de materia, compartidas en el equipo docente.

Universitat de Barcelona, enero de 2009.

Miquel Martínez

Director del Instituto de Ciencias de la Educación

RESUMEN

Gran parte de la evaluación de la docencia se centra en lo que los docentes hacen en clase. Este artículo, en cambio, se ocupa de evaluar cómo se realiza la evaluación y cómo ésta influye en la forma de aprender y estudiar fuera de clase. Generalmente se asume que la manera de evaluar determina poderosamente qué, cómo y cuánto estudian los alumnos. Este artículo propone un conjunto de «condiciones bajo las cuales la evaluación favorece el aprendizaje» y las argumenta con referencias teóricas, evidencias empíricas y experiencias prácticas. Estas condiciones ofrecen, además, un marco de referencia a los docentes dispuestos a revisar la efectividad de su propia práctica evaluativa.

INTRODUCCIÓN

Cuando la docencia universitaria aparece en los titulares de los periódicos es casi siempre por cuestiones de evaluación: son noticia una suelta bajada de nivel, la detección de casos de plagio, errores en los exámenes, etc. La actuación más reciente de la Agencia de la Calidad de la Educación Superior del Reino Unido (QAA) se ha centrado en los resultados de aprendizaje y en cómo evaluarlos, en especificar los niveles estándar y el papel de los examinadores externos para asegurar esos niveles, mientras que las medidas tomadas como estrategias institucionales de enseñanza y aprendizaje aplicadas a la evaluación se centran, casi siempre, en sintonizar los resultados de aprendizaje con la evaluación y en especificar los criterios de esa evaluación. El centro de interés de todos, desde los medios de comunicación, pasando por los responsables de calidad, hasta las mismas instituciones, se sitúa en la evaluación concebida como sistema de medición. Este artículo no trata en absoluto de sistemas de medición, trata del aprendizaje. Los sistemas de evaluación más seguros, fiables y rigurosos a menudo van de la mano de aprendizajes superficiales y estériles y conducen a resultados de aprendizaje de corta duración –de hecho a menudo son la vía directa a este tipo de aprendizaje superficial–. No estamos defendiendo una evaluación poco fiable, pero sí que las formas de evaluación deben diseñarse primordialmente para fomentar un aprendizaje fructífero a largo plazo y que, de la fiabilidad, hay que ocuparse después. Los niveles subirán gracias a la mejora del aprendizaje de los estudiantes y no tanto por la mejora de la medición de un aprendizaje limitado. Este artículo trata de cómo diseñar una evaluación que fomente el aprendizaje en profundidad. Los estudios de caso publicados sobre esta cuestión describen métodos de evaluación particulares –tácticas, si se quiere–. Una orientación sobre cómo implementar un amplio espectro de tácticas evaluativas puede encontrarse en otras publicaciones (p. ej. Gibbs, 1995). Este artículo trata de la estrategia –de las funciones de la evaluación (Gibbs, 1999)–, que permite a los docentes seleccionar las tácticas evaluativas más apropiadas. Demostraremos que la evaluación es más efectiva en términos de aprendizaje si se cumplen una serie de condiciones. En este artículo se examinará la naturaleza de estas condiciones.

LA INFLUENCIA DETERMINANTE DE LA EVALUACIÓN

A principios de los años setenta, investigadores de ambos lados del Atlántico (Snyder, 1971; Miller y Parlett, 1974) se dedicaron a investigar cómo era el aprendizaje de los estudiantes en algunas universidades de prestigio. Lo que encontraron fue que, sorprendentemente, no era la docencia lo que más influía en los estudiantes, sino la evaluación. Los estudiantes describían todos los aspectos de su actividad como estudiantes (a qué clases asistían, cuánto trabajaban y cómo les iba en los estudios) como si estuvieran completamente determinados por la manera en que percibían las demandas del sistema de evaluación. Derek Rowntree llegó a afirmar que «si queremos descubrir la verdad sobre el sistema educativo, primero tenemos que echar un vistazo a sus procedimientos de evaluación» (Rowntree, 1987, p. 1). Los estudios de Snyder y Miller y Parlett fueron más lejos y pusieron de manifiesto cómo los estudiantes responden a esos procedimientos de evaluación. Más recientemente, estudios cualitativos han puesto el énfasis en lo importante que es entender de qué manera responden los estudiantes a las innovaciones en evaluación (Sambell y McDowell, 1998).

El trabajo de Snyder dió origen a la noción del «currículum oculto», distinto al currículum formal descrito en la documentación de la asignatura, que los estudiantes tienen que ir descubriendo y prestándole atención si quieren tener éxito:

«Desde el principio, tuve la impresión de que todo consistía en una especie de ejercicio de administración de tiempo... Tienes que entresacar lo que es realmente importante en cada asignatura... materialmente no puedes hacerlo todo. Me di cuenta de que, si eres bueno entresacando lo que es importante, puedes hacerlo lo bastante bien como para sacar adelante todas las asignaturas sin problemas.» (Snyder, 1971, pp. 62-63)

Una vez averiguado en qué consiste este currículum oculto, los estudiantes pueden distribuir su esfuerzo con gran eficiencia:

«Ahora no me preocupo de hacer las tareas. Me planteo las asignaturas de manera que pueda sacar un sobresaliente de la manera más fácil, y

es increíble el poco trabajo que hay que hacer si realmente no te gusta la asignatura.» (Snyder, *ibid.* p. 50)

Miller y Parlett se centraron en investigar hasta qué punto los estudiantes actuaban en función de los indicios de qué era recompensado por el sistema de evaluación. Identificaron distintos tipos de estudiantes: el «buscador de indicios» que se esmeraba especialmente en sacarle al docente qué iba a salir en el examen y cuáles eran sus preferencias personales; el «consciente de los indicios» que escuchaba y prestaba atención a los consejos dados por los docentes sobre qué era importante y qué no, y el «sordo a los indicios», a quien cualquier orientación le entraba por un oído y le salía por el otro. Uno de estos estudiantes «buscadores de indicios» describe el proceso de descubrir las preguntas del examen:

«Estoy convencido de que se trata de un juego, del juego de examinarse. Ciertos datos, por ejemplo, no los estudias, y no vas a todas las clases, vas y te miras los exámenes y dices ‘por lo que veo había cuatro preguntas sobre un tema concreto ese año, el año pasado el profesor dijo que el examen sería más o menos parecido al del año anterior’, así que reduces un buen trozo de la asignatura inmediatamente...» (Miller y Parlett, 1974, p. 60)

En contraste, los estudiantes descritos como «sordos a los indicios» manifestaron lo siguiente:

«No selecciono cuestiones para preparar el examen –no me siento seguro si sólo me limito a ciertos temas.»
«Intentaré estudiármelo todo...» (Miller y Parlett, 1974, p. 63)

Miller y Parlett fueron capaces de predecir con gran precisión qué estudiantes obtendrían mejores resultados de graduación:

«La gente que era “consciente de los indicios” tendía a llegar a ser *upper seconds* y los “sordos a los indicios” se clasificaban como *lower seconds*.»¹ (Miller y Parlett, 1974, p. 55)

1. En las universidades del Reino Unido, según los resultados obtenidos, la graduación se califica de primera clase (70%+), segunda superior (60% - 69%) también llamada «2:1» o *Upper Second class*, segunda inferior (50% - 59%) o «2:2» *Lower Second class*, tercera clase (40% - 49%) o «3» *Third class*, o insuficiente (30% - 39%) o muy deficiente (0% - 29%).

Muchos estudiantes son perfectamente capaces de distinguir entre aquello a lo que han de prestar atención por cuestiones de evaluación y lo que es efectivo para un aprendizaje en profundidad, tal como explica este estudiante de Oceanografía:

«Si estás bajo mucha presión, entonces sólo te centras en sacarte la asignatura. Lo sé por propia experiencia, una mala experiencia. En una materia en la que no era muy bueno intenté entender los temas y suspendí el examen. Cuando volví a hacer el examen, me centré simplemente en conseguir pasar el examen. Conseguí un 96% y el profe no podía entender cómo había podido suspender la primera vez. Se lo dije: Esta vez me he centrado únicamente en conseguir pasar el examen y no en entender la materia. Sigo sin entender la materia, así que en cierta manera he fallado en lo que me proponía.» (Gibbs, 1992, p. 101)

Independientemente de que esté claramente especificado en la documentación de la asignatura qué se pretende evaluar, los estudiantes averiguan por sí mismos qué es lo que cuenta –o, por lo menos, lo que ellos creen que cuenta– y dirigen sus esfuerzos de manera acorde. Son estratégicos en su uso del tiempo y «selectivamente negligentes» evitando contenidos que creen que no serán evaluados. Se ha afirmado que los estudiantes, desde los setenta, se han vuelto más estratégicos en su uso del tiempo y de sus energías y que están cada vez más influidos, en lugar de menos, por las demandas percibidas del sistema de evaluación, en el sentido de que planifican más su paso por los estudios (MacFarlane, 1992).

LA IMPORTANCIA DE LAS ACTIVIDADES Y TRABAJOS DE CURSO

Los estudiantes tienden a obtener mejores calificaciones en los trabajos que en los exámenes (James y Fleming, 2004-2005). Chansarkar y Raut-Roy (1987) estudiaron los efectos de la combinación de formas variadas de trabajos y actividades con exámenes. Encontraron que todas las combinaciones de trabajos y actividades de distintos tipos con exámenes producían una media mejor en los índices de resultados que si se realizaban únicamente exámenes, la media de las calificaciones era hasta un 12% más alta.

Gibbs y Lucas (1997) analizaron las calificaciones de 1712 módulos de la Universidad Politécnica de Oxford. En los módulos diseñados al 100% con actividades y trabajos, las calificaciones medias eran un 3,5% superiores a las obtenidas en los módulos donde únicamente se realizaban exámenes. Hallaron una correlación positiva entre la proporción de actividades y trabajos existente en un módulo y la media de las calificaciones ($r = +0,36$, $p < 0,0001$). Bridges *et al.* (2002) estudiaron las diferencias en las calificaciones obtenidas en actividades y trabajos y en exámenes en seis titulaciones en cuatro universidades. Encontraron que las calificaciones obtenidas en actividades y trabajos de curso eran más altas en un tercio de la clasificación por titulaciones en Estudios Ingleses y en Historia (similar a lo encontrado en la Politécnica de Oxford) y más alta en dos tercios en Biología, Empresariales, Ingeniería de Sistemas y Derecho.

Los estudiantes, de hecho, prefieren realizar actividades y trabajos de curso. Starr (1970) ya indicaba que el 90% de los estudiantes de cuatro departamentos preferían que la mitad o más de sus calificaciones procedieran de actividades y trabajos de curso y el 56% prefería que todas sus calificaciones procedieran de trabajos de curso. Los estudiantes consideran que los trabajos y actividades de curso son más justos que los exámenes porque miden un espectro más amplio de habilidades que los exámenes y porque les permiten organizar sus propias formas de trabajo en mayor medida (Kniveton, 1996).

Una media más alta de las calificaciones o las preferencias de los estudiantes no serían argumentos suficientes si las actividades y los trabajos de curso fueran inherentemente menos válidos como forma de evaluación acreditativa. Pero no lo son. En primer lugar, los exámenes son muy pobres predictores de cualquier actuación subsiguiente, como la inserción y el éxito profesional. Un análisis de 150 estudios centrados en la relación entre resultados de exámenes y una amplia gama de indicadores de logros en la vida adulta pone de manifiesto que la relación es, en el mejor de los casos, tenue (Baird 1985). Por ejemplo, los resultados en el primer grado explican menos de un 10% de la varianza en la actuación en los postgraduados (Warren, 1971).

En segundo lugar, las calificaciones de las actividades y trabajos de curso son un mejor predictor de aprendizaje en profundidad y a largo plazo de los contenidos de las asignaturas que los exámenes. Conway *et al.* (1992) estudiaron los resultados de los estudiantes de Psicología en una variedad de tests para evaluar su capacidad de entender y recordar los contenidos de una asignatura de psicología cognitiva a la que habían asistido hacía muchos años. Encontraron que las calificaciones obtenidas en actividades y trabajos de curso realizados hasta 13 años antes correlacionaban con los resultados de estos tests, mientras que las calificaciones de los exámenes no. Podemos suponer que el tipo de aprendizaje implicado en las actividades y los trabajos de curso tiene efectos a largo plazo mientras que el tipo de aprendizaje implicado en el estudio para un examen no lo tiene. Los estudios sobre aprendizaje superficial o en profundidad ponen de manifiesto resultados similares: que los efectos positivos en los resultados de los exámenes de los estudiantes que optaban por un aprendizaje superficial al preparar los exámenes se mantenían sólo a corto plazo (Marton y Wenestam, 1978).

En tercer lugar, estudios experimentales en los cuales los estudiantes no se habían preparado ni para exámenes ni para realizar actividades y trabajos de curso han demostrado que la calidad de su aprendizaje es más alta en las asignaturas basadas en actividades y trabajos. Por ejemplo, Tynjala (1998) ha comparado dos grupos de estudiantes: el primer grupo estudiaba mediante clases expositivas convencionales, siguiendo un manual y realizando un examen; el segundo grupo estudiaba a través de actividades basadas en el manual, con debates con los otros

estudiantes sobre las actividades y elaborando un trabajo escrito que evaluaba el profesor. Este segundo grupo también realizó el examen final para permitir la comparación con el primer grupo a pesar de no haber estudiado para el examen. Se encontró que el segundo grupo ponía más énfasis en la elaboración conceptual y había desarrollado concepciones más complejas de qué significa aprender (v. Säljö, 1982). En las respuestas de sus exámenes utilizaban más comparaciones, más valoraciones y construcciones más complejas en los términos de la taxonomía SOLO sobre resultados de aprendizaje (Biggs y Collis, 1982). Estos resultados, alcanzados con menos horas de docencia, se interpretaron como productos mejor construidos en función de los requerimientos de evaluación del segundo grupo.

Existe la opinión generalizada entre los docentes universitarios de que si se suprimen actividades y trabajos de un módulo debido a restricciones de medios, los estudiantes simplemente dejan de hacer el estudio asociado; por ejemplo: los estudiantes raramente escribirán trabajos que no vayan a ser evaluados. Se argumenta que hay que evaluarlo todo para conseguir que los estudiantes pongan el tiempo y el esfuerzo para realizarlo. Sin embargo, las actividades y trabajos de curso no tienen por qué ser calificadas para generar el aprendizaje necesario. Forbes y Spence (1991) describen la investigación llevada a cabo en las prácticas evaluativas de una asignatura de Ingeniería en la universidad de Strathclyde. Cuando los docentes dejaron de poner notas a los ejercicios de problemas semanales –simplemente porque estaban desbordados– los estudiantes, en efecto, dejaron de hacer los problemas y las notas de los exámenes descendieron a consecuencia de ello. Pero cuando los docentes introdujeron una evaluación entre iguales periódica de los ejercicios de problemas –como actividad obligatoria del curso pero sin contribuir a la nota final– los resultados de los exámenes subieron espectacularmente hasta un nivel muy por encima del alcanzado con anterioridad, cuando los docentes puntuaban los ejercicios. Lo que hizo que aprendieran fue la calidad de la implicación de los estudiantes en las actividades de aprendizaje, no que los docentes realizaran gran cantidad de correcciones. El elemento crucial a la hora de diseñar sistemas de evaluación es generar la implicación de los estudiantes en las actividades de aprendizaje sin generar montañas de papel para corregir.

EL DECLIVE DE LA EVALUACIÓN FORMATIVA

Lo que ha caracterizado tradicionalmente la docencia universitaria en el Reino Unido ha sido ofrecer *feedback* personalizado y detallado sobre las actividades y trabajos de curso de manera frecuente. El arquetipo ha sido el de las Universidades de Oxford y Cambridge, donde los estudiantes elaboraban un ensayo semanalmente y se lo leían a su tutor durante la tutoría individual en la que recibían, de inmediato y de forma detallada, los comentarios críticos suscitados por la elaboración conceptual expuesta en el trabajo leído. Esto era prácticamente toda la enseñanza que muchos de sus estudiantes recibían: enseñar significaba dar *feedback* sobre los ensayos. Esta evaluación formativa estaba bastante separada de la evaluación sumativa y en Oxford y Cambridge la única evaluación sumativa consistía a menudo en unos exámenes finales al término de los tres años de estudios que habían incluido la evaluación formativa semanal.

Muy pocas instituciones han sido capaces de llegar a la cantidad o calidad del *feedback* ofrecido por Oxford y Cambridge, pero la mayoría asumía que actividades y trabajos frecuentes junto a *feedback* detallado (por escrito) era básico para el aprendizaje de los estudiantes. Hasta hace relativamente poco, por ejemplo, muchas titulaciones de ciencias incluían ejercicios semanales de problemas e informes de laboratorio, todos ellos evaluados y puntuados por los docentes y devueltos a los estudiantes en el plazo de una semana. En la mayor parte de las modalidades de educación a distancia, el *feedback* sobre las actividades frecuentes es el componente interactivo principal de la docencia y la Open University ha puesto un gran énfasis en actividades frecuentes, en formar y remunerar a tutores para ofrecer *feedback* extenso y detallado, y en realizar el control y seguimiento que garantice la calidad de ese *feedback*. Para algunos estudiantes de la Open University, éste es el único contacto que tienen con su tutor. Pueden pasar sin mucho contacto presencial, o incluso sin ninguno, pero no pueden pasar sin *feedback* periódico sobre las actividades y trabajos realizados.

En las universidades convencionales las limitaciones de recursos han traído consigo una reducción de la frecuencia de las actividades y tra-

bajos, de la calidad y cantidad de *feedback* ofrecido y un aumento sobre el tiempo de respuesta más adecuado para ofrecer el *feedback*. La modularización ha tendido a acortar el tiempo de las asignaturas y ha reducido la escala temporal en la que era posible proponer actividades y dar *feedback* y ha incrementado el número de exámenes. En algunos cursos se han abandonado completamente las actividades cuyo fin era la evaluación formativa. En otros quizá se mantenga una única actividad o trabajo pero con *feedback* muy retardado, a veces incluso después del examen. Al mismo tiempo la diversidad de estudiantes ha crecido enormemente, de manera que las presuposiciones sobre el nivel de complejidad de los conocimientos previos, técnicas y hábitos de estudio, la propia concepción de qué es el aprendizaje (Säljö, 1982) o la concepción de qué es el conocimiento (Perry, 1970) que aportan los estudiantes difícilmente son acertados. Más bien parece que hay que incrementar la orientación y ofrecer una guía a los estudiantes que necesitan más experiencia para realizar las actividades y trabajos y más *feedback* sobre su aprendizaje, en lugar de disminuirlo. En la educación a distancia, por el contrario, precisamente porque las actividades periódicas y el *feedback* detallado son considerados como elementos cruciales, éstos se han mantenido ampliamente; como consecuencia, los estudiantes de la Open University hoy puede que reciban 50 veces más *feedback* sobre sus actividades de aprendizaje a lo largo de sus estudios que los estudiantes de las universidades convencionales.

LA EFECTIVIDAD DEL *FEEDBACK*

En una revisión en profundidad de 87 metaanálisis de investigaciones sobre qué es realmente significativo para conseguir buenos resultados de aprendizaje, Hattie (1987) encontró que el factor individual más poderoso era el *feedback*. La revisión de las investigaciones sobre evaluación formativa que Black y William (1998) ofrecen destaca en términos parecidos los efectos positivos sobre el aprendizaje, consistentes y a largo plazo, que tiene el *feedback* en comparación con otros aspectos de la docencia. Ha habido muchos intentos de comprender tanto la naturaleza de este impacto como de explotar su potencialidad a través de la innovación, por lo menos en las escuelas, como consecuencia de esta evidencia incontrovertible.

En educación superior, el *feedback* individual ofrecido a los estudiantes en clase parece que ha descendido de manera significativa a medida que el tamaño de los grupos aumentaba, aunque no ha sido documentado. Los comentarios escritos sobre las actividades y trabajos, sin embargo, siguen siendo uno de los mayores componentes de la dedicación del profesorado en educación superior. Al aumentar el tamaño de los grupos se han ido introduciendo algunas medidas de economía de escala en la docencia (incrementando simplemente el número de estudiantes por aula), pero las economías de escala son difíciles de realizar con éxito en la evaluación: los costes de evaluación aumentan en proporción directa al número de estudiantes. Como consecuencia, los costes de evaluación pueden acabar acaparando los costes de docencia y los docentes pueden encontrarse dedicando la mayor parte de su tiempo a corregir y a poner notas. ¿Vale la pena todo este esfuerzo?

En el Course Experience Questionnaire (Ramsden, 1991), usado ampliamente en Australia y otros países para evaluar la calidad de los cursos, el ítem del cuestionario que distingue más claramente los mejores y los peores cursos es «el personal docente aquí ofrece *feedback* útil sobre cómo vas» (Ramsden, 1992, p. 107). Esto no significa que los docentes universitarios ofrezcan efectivamente *feedback* útil, sólo quiere decir que lo ofrezcan o no, eso es lo más significativo de todo lo que hacen. En realidad, ¿hasta qué punto es efectivo ofrecer *feedback*?

MacLellen (2001) preguntó a 130 estudiantes y a 80 docentes de la Universidad de Strathclyde sobre su percepción de la evaluación. Entre las 40 preguntas planteadas, 4 eran sobre *feedback* y éstas revelaron grandes discrepancias entre estudiantes y docentes. Mientras que la mayor parte de los docentes respondieron que el *feedback* detallado es útil «con frecuencia», ayuda al estudiante a entender «con frecuencia» y contribuye al aprendizaje «con frecuencia», la mayoría de estudiantes respondieron que el *feedback* era sólo «a veces» útil en esos casos. El 30% de los estudiantes indicaron que el *feedback* «nunca» les ayuda a entender. Mientras que el 63% de los docentes respondieron que el *feedback* incita «con frecuencia» a consultar al tutor, sólo el 2% de los estudiantes respondieron en el mismo sentido y el 50% respondió que el *feedback* «nunca» incita a consultar al tutor.

Puede que aquí haya tal problema con la cantidad y calidad del *feedback* que, efectivamente, invalide la efectividad del *feedback* ofrecido a los estudiantes –después de todo, los docentes están bajo una enorme presión de tiempo y de trabajo y es difícil en estas circunstancias ofrecer *feedback* extenso, detallado y útil–. Pero hay también otros problemas. Leer las investigaciones sobre qué hacen los estudiantes con el *feedback* causa depresión. A menudo el *feedback* ni siquiera se lee (Hounsell, 1987) o no se entiende (Lea y Street, 1998). Wotjas (1998) indica:

«Algunos estudiantes, si les disgusta la nota, tiran el *feedback* a la papelera, mientras que a otros parece que sólo les interesa el resultado final y ni siquiera pasan a recoger el trabajo corregido.»

Hay también otro problema relacionado tanto con las notas como con el *feedback* ofrecido. El estudiante percibe probablemente la calificación como indicador de su habilidad personal o incluso como una valoración de su persona, porque en general se trata de una evaluación normativa que indica dónde se sitúa uno en relación a los otros. Una calificación baja puede dañar la «autoeficacia» de un estudiante, es decir, la sensación que tiene de ser eficaz. Yorke (2001) profundiza en las maneras positivas y negativas en que la evaluación formativa puede afectar la retención de estudiantes y pone énfasis en el papel que ésta puede tener en la «integración académica» (Tinto, 1993). En contraste, el *feedback* es más probable que sea percibido por su parte como comentario sobre qué se

ha aprendido. Cuando no hay calificaciones, se ha demostrado que los estudiantes leen el *feedback* con mucha mayor atención (Black y William, 1998) y lo usan para guiar su aprendizaje. A la luz de las evidencias de la investigación (en el ámbito escolar), algunas escuelas han adoptado la política de ofrecer únicamente *feedback* sobre las actividades o trabajos realizados y no puntuarlos. El sistema de «evaluación como aprendizaje» del Alverno College es probablemente el ejemplo más conocido en Educación Superior de evaluación «con menos notas».

El cuadro no es demasiado prometedor. De una forma u otra, la evaluación parece ser enormemente costosa, desagrada tanto a estudiantes como a profesores y es ampliamente ineficaz para favorecer el aprendizaje. Ante semejante situación, nos proponemos plantear y justificar a continuación una serie de «condiciones bajo las cuales la evaluación favorece el aprendizaje». Las evidencias son en pocos casos suficientemente concluyentes como para argumentar que si la evaluación cumple con estas condiciones entonces el aprendizaje será forzosamente más efectivo. Por lo tanto, son ofrecidas más bien como un conjunto de orientaciones plausibles.

Éste no constituye el primer intento de identificar unas «condiciones» de este tipo, pero sí lo es en el contexto de la Educación Superior. La investigación en el ámbito escolar ya identificó los efectos de la evaluación formativa y los enumeró como puede verse en la siguiente lista basada en Gagne (1977):

1. Reactivar o consolidar habilidades o conocimientos previos necesarios antes de introducir nueva materia.
2. Centrar la atención en los aspectos más importantes de la materia.
3. Incentivar las estrategias de aprendizaje activo.
4. Ofrecer oportunidades a los estudiantes para practicar sus habilidades y consolidar el aprendizaje.
5. Dar a conocer los resultados y ofrecer *feedback* correctivo.
6. Ayudar a los estudiantes a controlar su propio progreso y desarrollar las capacidades de autoevaluación.
7. Orientar en la toma de decisiones sobre las siguientes actividades de instrucción o aprendizaje para aumentar su dominio.
8. Ayudar a los estudiantes a sentir que han alcanzado un objetivo.

Las condiciones esbozadas aquí ejercen dos tipos de influencia de naturaleza relativamente diferente:

- La influencia, a través del diseño de los sistemas de evaluación y de las actividades y trabajos, sobre cuánto estudian, qué estudian y de qué manera se implican los estudiantes.
- La influencia, a través del *feedback*, sobre el aprendizaje.

CÓMO INFLUIR A TRAVÉS DE LA EVALUACIÓN SOBRE EL VOLUMEN, LA SELECCIÓN Y CALIDAD DE LO ESTUDIADO

Condición I

Ofrecer a los estudiantes suficientes actividades evaluables para ocupar el tiempo de estudio necesario y suficiente

Esta cuestión está relacionada con cuánto tiempo y esfuerzo dedican los estudiantes al objetivo de aprendizaje –el principio de «tiempo dedicado a una tarea» (Chickering y Gamson, 1987), según el cual si los estudiantes no dedican suficiente tiempo a algo no van a aprenderlo. Berliner (1984), resumiendo la investigación realizada sobre «tiempo dedicado a una tarea», concluyó que hay evidencias empíricas más que suficientes para establecer una relación directa entre tiempo dedicado a las asignaturas, gestión del tiempo del estudiante y tiempo real del estudiante dedicado a la realización de la tarea, por un lado, y los resultados de aprendizaje del estudiante, por el otro.

La relación entre esfuerzo y calificación no es siempre directa. Kember *et al.* (1996) encontraron que las percepciones de los estudiantes sobre su esfuerzo dependían más de su motivación que del número de horas que realmente dedicaban y que se daba el caso de estudiantes que «echaban muchas horas» improductivas, especialmente si adoptaban un estilo de aprendizaje superficial en sus estudios. Algunos tipos de evaluación pueden generar muchas horas de memorización ineficaz.

El diseño de las asignaturas en las universidades del Reino Unido tiene en cuenta un número determinado de horas de aprendizaje relacionado con el número de créditos de la asignatura. Normalmente se espera que los estudiantes dediquen al estudio entre una y cuatro horas por cada hora pasada en clase (dependiendo en gran medida de la disciplina en cuestión). Innis (1996) encontró que los estudiantes de la Leeds Metropolitan University dedicaban entre 1,4 y 3,0 horas de estudio por cada hora de clase. Cuánto de este tiempo «fuera de clase» se dedica realmente a estudiar viene determinado en gran parte por lo que exige la forma

de evaluación. En Estados Unidos, los estudiantes universitarios dedican al estudio, de media, menos de la mitad del tiempo que esperan sus profesores: entre 0,3 y 1,0 horas de estudio fuera de clase, mientras los docentes esperan, de media, 2,1 horas por cada hora de clase (Moffat, 1989; Hutchings *et al.* 1991; Gardiner, 1997; Brittingham, 1988). En Estados Unidos, cuando se ha intentado mejorar los resultados de los estudiantes a través de la evaluación, el énfasis se ha puesto en la «evaluación en clase» –es decir, en actividades que se realizan en clase para evaluar a los estudiantes, generalmente mediante tests– y esta información suministrada por la evaluación se usa para orientar tanto a los estudiantes como a los docentes (Angelo y Cross, 1993). Esta atención especial a las actividades dentro de clase puede ser interpretada como el reconocimiento del fracaso en generar gran parte del aprendizaje fuera de clase a través del tipo de evaluación que usan. Los estudios realizados a través de diarios de aprendizaje (p. ej. Innis, 1996) muestran que los estudiantes en el Reino Unido dedican su tiempo mayormente a las actividades que serán evaluadas y que esto, a la larga, se va convirtiendo en un centro de atención cada vez más restringido a medida que van adquiriendo experiencia como estudiantes, hasta llegar a dedicar, hacia el tercer año, únicamente un 5% de su tiempo a actividades de estudio disociadas de una evaluación.

Las áreas disciplinares con actividades que se evalúan con menor frecuencia (por ejemplo, aquellas basadas en textos) cuentan con los estudiantes que dedican una menor cantidad de horas (Vos, 1991). Las áreas de ciencias y tecnología, que generan una dedicación al estudio mayor, tienden a programar actividades que se evalúan con mayor frecuencia (y que son también más limitadas), como hojas de ejercicios de problemas e informes de laboratorio.

Los estudios realizados para indagar cómo influye el trabajo remunerado ejercido en paralelo a la dedicación al estudio a tiempo completo muestran que esos estudiantes dedican menos horas (Curtis y Shami, 2002) y que sus resultados son significativamente menos buenos (Paton-Salzberg y Lindsay, 1993). Estos estudios muestran que más de las tres cuartas partes de los estudiantes a tiempo completo trabajan durante el periodo lectivo y que distribuyen con toda probabilidad sus horas de estudio reducidas de forma especialmente estratégica en rela-

ción a los requerimientos de la evaluación. Indican que reducen lecturas y otras actividades de estudio fuera de clase.

Proponer actividades y trabajos que se evalúen no es la única forma de atraer el tiempo y el esfuerzo de los estudiantes. La forma convencional de hacerlo es realizar exámenes de una selección imprevisible de contenidos de la asignatura sin avisar, de manera que para un estudiante ignorarlo todo se convierta en una actividad de alto riesgo. La calidad, la cantidad y la distribución del esfuerzo de estudio conseguido de esta manera son impredecibles y probablemente varían en función de la percepción que tienen los estudiantes de las demandas probables del examen y de los riesgos implicados.

También se puede conseguir que los estudiantes dediquen el tiempo y el esfuerzo necesario a través de la presión social, por ejemplo apelando:

- al ridículo potencial ante los compañeros si éstos pueden contemplar la baja calidad del trabajo realizado, como sucede al evaluar exposiciones orales en un seminario o si los informes de laboratorio escritos se muestran públicamente en forma de póster;
- a la censura potencial de los compañeros si uno de los integrantes de un grupo de trabajo (que será evaluado) no realiza su parte correctamente.

Condición 2

Ocupar a los estudiantes con actividades evaluables que les orienten sobre cómo distribuir apropiadamente el tiempo y el esfuerzo en función de los aspectos más importantes de la asignatura

Esta condición atañe a la dirección y calidad del esfuerzo, hacia dónde debe orientarse y hasta qué punto. En general, los estudiantes distribuyen su tiempo de manera irregular entre las asignaturas, a menudo de forma concentrada en los temas relacionados con la evaluación y nada más. Si dibujaran un gráfico con el esfuerzo semanal dedicado a lo largo de todas las semanas del curso a una asignatura que tuviera asignada una secuencia de actividades, éste se asemejaría más a los Al-

pes que a Holanda. Los exámenes pueden tener el efecto de concentrar el estudio en cortos pero intensos períodos de actividades de estudio al final del curso (como, por ejemplo, revisar apuntes) hasta muchas semanas después de haber terminado la actividad docente. Actividades de clase frecuentes (como las hojas de ejercicios o problemas) o tests (como los exámenes computerizados) pueden distribuir el esfuerzo de los estudiantes a lo largo de la asignatura, a menudo en una distribución semanal, mientras que actividades poco frecuentes (como los trabajos escritos largos tipo ensayo) pueden dar como resultado un estudio intensivo de una semana o dos justo antes de la fecha de entrega, al tiempo que temas no cubiertos por las actividades propuestas pueden ser totalmente ignorados.

De hecho, sabemos muy poco sobre cómo distribuyen su esfuerzo los estudiantes, y los docentes universitarios tienden a saber muy poco sobre qué hacen sus estudiantes y cuándo.

Condición 3

Afrontar una tarea a evaluar implica realizar unas actividades de aprendizaje productivas del tipo más apropiado

Esta cuestión atañe a los diversos tipos de estudio y de actividades de aprendizaje implicados en la manera de afrontar una tarea o de prepararse para un test. Algunas formas de evaluación conllevan actividades de aprendizaje inútiles e inapropiadas, a pesar de que puedan ir asociadas a calificaciones fiables. Estudiar para un examen de respuesta múltiple puede llevar a los estudiantes a un tipo de aprendizaje superficial (Scouler y Prosser, 1994; Tang, 1994; Scouler, 1998), al igual que los exámenes, aunque el estilo de estudio pueda influir tanto como la forma del test. Los estudiantes pueden optar por un aprendizaje en profundidad al prepararse para los tests de respuesta múltiple y adoptar estrategias de estudio efectivas, a pesar de que el test sólo plantee requerimientos de bajo nivel, y Macdonald (2002) ha indicado que por lo menos algunos estudiantes adoptaban un estilo de aprendizaje en profundidad cuando se preparaban para los exámenes y aprendían de forma efectiva, como resultado de integrar el contenido implicado en la preparación.

La evaluación falla muchas veces simplemente porque no consigue ocupar a los estudiantes realizando los tipos de aprendizaje apropiados. Entregar un informe, tras realizar en el laboratorio un procedimiento diseñado por el profesor, difícilmente ayudará al estudiante a aprender a diseñar experimentos. Probablemente la única forma de aprender a resolver problemas sea resolviendo un montón de problemas. Probablemente la única forma de adquirir una cierta fluidez en el lenguaje de la disciplina sea practicar una y otra vez el uso de ese lenguaje, por ejemplo, a través de la escritura. Las tareas y trabajos son la forma principal de generar este tipo de prácticas. Los estudiantes raramente se implicarán en serio en prácticas tan exigentes si éstas no cuentan para la evaluación de rendimiento o no son, por lo menos, un requerimiento dentro de las condiciones de evaluación. Parece improbable que los estudiantes lleguen a escribir trabajos del tipo ensayo y que aprendan con ello lo que eso implica, si no están obligados:

«En cierto sentido, es sólo trabajo. Todos esos ensayos, y las lecturas son lo peor, son en realidad sólo un montón de horas.» (Estudiante de Historia) (Hounsell, 1987)

Algunas formas de evaluación pueden orientar mal el esfuerzo de los estudiantes. Snyder (1971) ya describió cómo estudiantes del Massachusetts Institut of Technology, a los que se les pedía ser imaginativos, abandonaban cualquier iniciativa al respecto al descubrir que las notas se derivaban principalmente de la pura memorización de materia para exámenes de respuesta múltiple. Algunas formas de evaluación crean actividades de aprendizaje apropiadas, como efecto colateral. Por ejemplo, pedir que se escriban trabajos de tipo ensayo puede contribuir en general a que se «lea más allá» y puede favorecer la elaboración de argumentos coherentes, de una manera que no se conseguiría pidiendo a los estudiantes que simplemente leyeran lo que se indica en la lista de lecturas. Si elimináramos el ensayo, la forma apropiada de estudio no llegaría a producirse nunca, ni siquiera en el hipotético caso de que se llegara a realizar una cantidad similar de lecturas de un tipo parecido. El producto, el ensayo, y las notas asociadas a él pueden ser en sí mismos menos importantes para el aprendizaje que el marco que crea la propia tarea y que conlleva actividades de aprendizaje de «leer más allá» y «construir argumentos». De forma similar, en el caso de los in-

formas de laboratorio o de un *brief* de diseño,² el producto puede ser menos importante que las particularidades de estudio a las que induce este tipo de tarea. Los proyectos en grupo pueden generar más debates entre los estudiantes y, por tanto, hacer que los individuos se enfrenten a puntos de vista alternativos y conozcan estándares de trabajo diferentes. La calidad del producto, fruto del trabajo en grupo (como pueda ser un informe) que se refleja en una nota puede ser menos importante que las cualidades de los procesos de aprendizaje creados por la tarea.

Pero los estudiantes pueden afrontar las tareas diseñadas como actividades de aprendizaje de tal manera que maximicen las notas que obtengan en lugar de maximizar el aprendizaje alcanzado al implicarse en la tarea. Esto puede significar dar una falsa apariencia de conocimiento, hacerse pasar por competente y conocedor, escondiendo de forma deliberada lo que no se entiende o se desconoce, diciendo a los profesores lo que quieren oír en lugar de manifestar lo que se piensa realmente, etc. Hasta cierto punto esto es el resultado de las preferencias del propio estudiante, pero las formas de evaluación, el régimen de notas y el uso del *feedback* pueden incluso ser más determinantes que las actitudes personales y llegar incluso a fomentar un tipo de comportamiento entre los estudiantes que minimice su aprendizaje. En el siguiente ejemplo un estudiante con una motivación intrínseca describe en un diario de aprendizaje la manera en que suele realizar las tareas en Ingeniería, pensando en obtener las mejores notas a expensas de la calidad de su aprendizaje:

«El profesor medio quiere ver los resultados correctos en un recuadro rojo al final de la hoja de ejercicios, a ser posible con tan pocas líneas de cálculo como sea posible –sobre todo esto–, no pongas ningún comentario. Lo odia. Piensa que estás intentando llenar la página con palabras para que parezca que tu trabajo es más largo. Tampoco dejes los errores aunque los hayas corregido. Si lo has hecho mal, todo a la papelera. Le gusta creer que has encontrado la solución a la primera. Si sigues haciendo fallos, eso quiere decir que no has estudiado lo bastante. No puedes volver a hacer un ejercicio unos meses más tarde porque tú

2. Documento escrito sobre el proyecto de diseño que integra lo que un cliente quiere de un diseñador para un trabajo en específico.

sólo tienes los resultados finales sin comentarios. Si quieres intentarlo, puede que vuelvas a hacer los mismos errores que has hecho antes porque no has guardado tus errores previos.» (Gibbs, 1992)

La influencia del *feedback* en el aprendizaje

«Saber qué sabes y qué no sabes permite centrar el aprendizaje. Los estudiantes necesitan *feedback* apropiado sobre cómo lo están haciendo para sacar provecho de las asignaturas. Al empezar, los estudiantes necesitan ayuda para evaluar qué conocimientos tienen y en qué nivel de competencia se encuentran. En las clases, los estudiantes necesitan tener con frecuencia la posibilidad de recibir sugerencias de mejora y ponerlas en práctica. En distintos momentos de sus estudios, y al acabar, los estudiantes necesitan tener la oportunidad de reflexionar sobre qué han aprendido, qué tienen que seguir aprendiendo y cómo van a autoevaluarse.» (Chickering y Gamson, 1987)

Tradicionalmente, el *feedback* se ha entendido como una cuestión de «corrección de errores» (Bruner, 1974) o «conocimiento de resultados» en relación al aprendizaje mismo; si a una estudiante se le dice que va por buen camino, entonces aprenderá. En este artículo nos ocupamos de cómo la provisión de *feedback* afecta el comportamiento de aprendizaje del estudiante, de cómo el *feedback* influye en la toma de decisiones del estudiante que implican, o no, un aprendizaje posterior.

Condición 4 Dar suficiente *feedback*, tanto con suficiente frecuencia como con el suficiente detalle

Esta cuestión está directamente relacionada con lo que convencionalmente se define como evaluación formativa: el impacto que tiene en el aprendizaje el *feedback* que se refiere al progreso realizado y que se ofrece por lo general inmediatamente después de haber realizado una tarea. La cantidad y el detalle del *feedback* varían considerablemente de una asignatura a otra –creemos que aquí la variación es todavía mayor que la variación existente en cantidad o calidad de la docencia entre asignaturas.

Este tipo de *feedback*, para ser útil, ha de ofrecerse con bastante regularidad y sobre porciones relativamente pequeñas de contenidos de la asignatura. Un único *feedback*, aunque sea detallado, sobre un trabajo extenso tipo ensayo o una tarea de diseño después de diez semanas de estudio difícilmente contribuirá a un mejor aprendizaje a lo largo de todo el curso. Los tests computerizados se han usado ampliamente para ofrecer, por lo menos, algo de *feedback* sobre el progreso realizado y en algunos casos, con la ayuda de software desarrollado específicamente para la evaluación, existe la posibilidad de ofrecer un «*feedback* de repaso» si la respuesta seleccionada era incorrecta. Cook (2001) encontró que las notas de los exámenes finales de los estudiantes están estrechamente relacionadas con el número (y, por lo tanto, la frecuencia) de las tareas de evaluación automatizada que los estudiantes habían realizado. La frecuencia y la rapidez de respuesta de este tipo de *feedback*, que puede ofrecerse de manera razonablemente económica, puede compensar su calidad, relativamente pobre, y su falta de individualización.

El *feedback* tiene que ser bastante específico para ser útil. La Open University forma a sus 7.500 tutores a tiempo parcial para que puedan dar un *feedback* detallado y suficientemente extenso. Cole *et al.* (1986) ofrecen una lista de las características del *feedback* efectivo en la educación a distancia y Roberts (1996) comprobó que las preferencias de los estudiantes, por lo que respecta al *feedback*, coinciden con esta lista. Las formas específicas de *feedback* que son efectivas varían de una disciplina a otra. Evidencias sobre las formas de *feedback* más efectivas en el aprendizaje de lenguas, por ejemplo, se encuentran resumidas en Hyland (2001). Tanto en Psicología (Stephenson *et al.* 1996) como en Matemáticas (Rice *et al.* 1994), los estudiantes han manifestado su preferencia por un *feedback* específico, detallado y motivador. Greer (2001) presenta un estudio que explica exactamente qué tipo de impacto tiene el *feedback* en el aprendizaje de contabilidad.

Gran parte del *feedback* ofrecido a los estudiantes en educación superior sería rechazado por los tutores de personal docente de la Open University (que realizan el seguimiento de la calificación realizada por los tutores) por ser completamente inadecuado y pondría en marcha mecanismos de garantía de la calidad y formación del profesorado.

Condición 5

Centrar el *feedback* en la actuación de los estudiantes, en su aprendizaje y en las acciones que están bajo su control, en lugar de centrarlo en los propios estudiantes o en sus características

La bibliografía existente sobre evaluación formativa distingue entre el *feedback* que le dice al estudiante que no tiene remedio, o que se encuentra dentro del 10% inferior de los estudiantes (en un nivel insuficiente, por ejemplo) y el *feedback* que le dice al estudiante dónde exactamente ha hecho algo mal y qué puede hacer para remediarlo. Calificaciones sin *feedback* pueden ser particularmente dañinas. Un *feedback* crítico centrado en las características personales puede desmotivar y puede afectar negativamente la «autoeficacia» o la sensación de competencia de los estudiantes. Esto es importante porque la «autoeficacia» está en estrecha correlación con la capacidad de esfuerzo y persistencia en las tareas (Schunk, 1984; 1985), es un buen predictor de logros académicos y va asociada a la adopción de un aprendizaje en profundidad (Thomas *et al.* 1987). Por el contrario, el *feedback* centrado en el contenido le da al estudiante opciones para actuar y está menos asociado a su ego –se trata de su actuación y no de él mismo–. Wootton (2002) ha escrito con pasión sobre el impacto negativo de la evaluación sobre los estudiantes «en riesgo» y se pregunta si el sistema existe para «fomentar el aprendizaje o para medir el fracaso».

Condición 6

Ofrecer el *feedback* a tiempo para que sea recibido cuando todavía le importa a los estudiantes y puedan utilizarlo en un aprendizaje posterior o puedan recibir ayuda a tiempo

Esta cuestión ya fue destacada en los «siete principios de buenas prácticas en educación superior a nivel de grado» (Chickering y Gamson, 1987; 1991). Se basa en una serie de estudios realizados sobre la temporización del *feedback* (para una síntesis, véase Dunkin, 1986; McKeachie *et al.* 1986). Un método docente que pone gran énfasis en el *feedback* inmediato en cada una de las fases de la progresión del estudiante a lo largo de las unidades de la asignatura, llamado *Personalised*

System of Instruction (PSI), ha demostrado, según indican numerosos estudios, que mejora la actuación del estudiante (Kulik *et al.* 1980).

Si los estudiantes no reciben *feedback* con la suficiente rapidez, ya habrán empezado a trabajar nuevos contenidos y el *feedback* resultará irrelevante para su estudio actual y será extremadamente improbable que genere una actividad apropiada de aprendizaje adicional que dé resultados. A causa de la presión por la falta de medios, se está dando el *feedback* con mayor lentitud y, dado que las asignaturas tienen ahora en el Reino Unido una duración menor, esto puede significar que el *feedback* sobre los trabajos de la asignatura se llegue a dar cuando ya se ha acabado ésta. Gran parte de este valioso y costoso *feedback* seguramente será en balde. Debería encontrarse un compromiso entre la celeridad y la calidad del *feedback* de un tipo como, por ejemplo, el *feedback* imperfecto de un estudiante que colabora en tareas de docencia, ofrecido casi de inmediato, y que puede tener un impacto mucho más positivo que el *feedback* de mucha mayor calidad de un tutor, ofrecido cuatro semanas más tarde.

Carroll (1995) describe unos «talleres de evaluación formativa» para grupos de 300 estudiantes de medicina que consistían en tests de respuesta múltiple seguidos de inmediato de una tutoría de «recuperación» sobre el tema. Este sistema no ofrecía un *feedback* individualizado, pero el *feedback* era inmediato y las sesiones del taller estaban programadas de tal manera que permitían a los estudiantes dedicar tiempo a estudiar más materiales antes de pasar a la siguiente sección del curso. El 85% de los estudiantes manifestaron que querían más sesiones de este tipo. Sly (1999) describe el impacto de los test «para practicar» la realización del examen subsiguiente. Los estudiantes tenían la opción de realizar «prácticas de test», con *feedback* computerizado, con suficiente tiempo de antelación antes del examen para permitirles utilizar el *feedback* cuando revisaran qué tenían que estudiar para mejorar en los puntos débiles. 197 estudiantes flojos escogieron realizar estas «prácticas de test» y todos mejoraron la puntuación obtenida en los exámenes de tal manera que superaron a 417 estudiantes, a priori más sólidos. Los efectos siguieron siendo evidentes en un examen posterior.

Condición 7

El *feedback* es apropiado en relación al propósito de la actividad y a los criterios de evaluación

Esta cuestión hace referencia a la relación entre el *feedback* y la finalidad de una actividad de aprendizaje y aquello que cuenta como una buena realización de esa actividad. El *feedback* puede tener diversas funciones. Por ejemplo, puede ser usado principalmente para:

- corregir errores,
- mejorar la comprensión a través de explicaciones,
- generar más aprendizaje al sugerir tareas de estudio específicas subsiguientes,
- desarrollar habilidades genéricas centrando la atención en el efecto del uso de esas habilidades en lugar de en el contenido,
- fomentar estrategias metacognitivas pidiendo a los estudiantes que se fijen en el proceso de aprendizaje implicado en la actividad propuesta y lo analicen,
- motivar a los estudiantes a seguir estudiando.

Cuál de todas estas funciones es la apropiada, depende en primer lugar de por qué se ha decidido realizar una actividad determinada. Por ejemplo, si la intención era ofrecer una única oportunidad para practicar la aplicación de un procedimiento o el uso de un algoritmo de una manera rigurosa o se trataba de ofrecer una oportunidad, entre muchas otras, de hacer prácticas en el uso de una competencia transversal, o se pretendía ofrecer una oportunidad compleja para reflexionar sobre el aprendizaje, o se quería poner un primer problema sencillo dentro de la asignatura para motivar al estudiante a no abandonar y continuar hasta el final.

Un estudio reciente en la Open University sugería que mantener la motivación era la cuestión más importante y que más influía en los estudiantes de nuevo ingreso al realizar la primera actividad de una asignatura (Gibbs y Simpson, 2002). Si un estudiante está buscando que le animen y sólo recibe correcciones a sus errores, es posible que no se favorezca su aprendizaje de la manera más efectiva.

Los estudiantes necesitan entender por qué han recibido la calificación o el comentario que han tenido y por qué no han recibido uno más alto (o más bajo). Los criterios tienen que ser explícitos, tienen que ser comprendidos por los estudiantes y quedar patentes en las calificaciones. A menudo, los criterios no van directamente ligados a parámetros y grados de consecución estándares y es difícil para un estudiante saber qué estándar es el que se espera de él o se consideraría inadecuado. Gran parte de la bibliografía sobre el uso de la autoevaluación y la evaluación entre iguales se ocupa de la fiabilidad de este tipo de evaluación y admite que la autoevaluación y la evaluación entre iguales es mayormente un mecanismo para ahorrar trabajo y personal. Pero el valor real seguramente hay que buscarlo en su contribución a que los estudiantes interioricen los estándares que se espera que alcancen, de manera que ellos puedan supervisarse y mejorar la calidad de sus propias actividades antes de entregarlas.

Los estudiantes necesitan entender los criterios para orientarse ellos mismos de manera apropiada en la tarea que se les propone. Penny y Grover (1996) han descrito de qué manera los estudiantes pueden llegar a malinterpretar los criterios que van a ser usados para evaluar su proyecto de investigación de final de carrera. Los estudiantes esperaban criterios relacionados con objetivos de bajo nivel, como el estilo y la presentación, mientras que los profesores ponían el énfasis en objetivos de alto nivel, como la comprensión teórica y conceptual. Creando oportunidades para dar *feedback* en múltiples momentos a lo largo del proyecto en curso se puede reorientar el esfuerzo de los estudiantes de manera apropiada (Carless, 2002).

La evaluación, por tanto, desempeña un papel importante a la hora de fijar los estándares a los que han de aspirar los estudiantes. Fijar las expectativas es uno de los «siete principios de buenas prácticas en educación superior a nivel de grado» (Chickering y Gamson, 1987). *Feedback*, respuestas modelo y muestras especialmente ejemplares (Orsmond *et al.* 2002) ayudan a establecer estas expectativas y la autoevaluación ayuda a los estudiantes a interiorizarlas.

Condición 8

El *feedback* es apropiado en relación a lo que entienden los estudiantes que tienen que hacer

Concepciones de la tarea que tienen los estudiantes

Los estudiantes tienen que entender qué tipo de tarea deberán realizar cuando se propone una actividad de aprendizaje y qué contará como «bueno» cuando lo intenten. Pueden tener concepciones erróneas o confusas, independientemente de la información inicial o el *feedback* que se les haya dado en el pasado, tal como puede verse en este ejemplo:

- «-¿Qué te parece que querría ver el tutor en este ensayo?
-Eh, ... bueno, voy un poco perdido. Sé que al tutor le gusta el trabajo conciso, que no le gusta que se generalice y que no le gusta que haya demasiados detalles, pero a pesar de todo, en general, yo creo que prefiero más detalles que generalizaciones. Y como era un tema tan general, pensé «socorro», no sé qué quiere que haga.» (Hounsell, 1987)

Sea cual sea el *feedback* que le dé el tutor al estudiante, éste será interpretado a partir de las ideas que el estudiante se haya formado sobre qué quiere realmente su tutor o en qué consiste realmente la tarea. Los estudiantes pueden tener grandes dificultades para comprender qué sentido tiene un ensayo (como forma escrita de comunicación cuando el único lector sabe más que ellos sobre el tema) o para qué sirve un informe de laboratorio (cuando ha sido escrito ya cientos de veces en el mismo formato exactamente) o para qué se ha propuesto una tarea de diseño (cuando sólo se evaluará el producto resultante y no el aprendizaje implicado en su creación). Muchas tareas en la universidad no tienen demasiado sentido para los estudiantes. Esto inevitablemente causa problemas cuando llega el momento de leer el *feedback* donde se les dice si han realizado de manera apropiada esa tarea incomprensible.

Concepciones del aprendizaje que tienen los estudiantes

Junto a la mencionada confusión de los estudiantes sobre qué está esperando realmente el tutor puede haber una concepción naíf del aprendizaje. Säljö (1982) encontró que los estudiantes conciben el aprendizaje de una de estas cinco maneras:

1. El aprendizaje como recepción pasiva de información.
2. El aprendizaje como memorización activa de información.
3. El aprendizaje como memorización activa de información o procedimientos para ser usados en algún momento en el futuro.
4. El aprendizaje como comprensión.
5. El aprendizaje como cambio de la realidad personal: percepción del mundo de manera diferente.

Un estudiante con las concepciones del aprendizaje 1, 2 o 3 puede tener problemas para interpretar un *feedback* que diga: «No hay suficiente discusión crítica», si le ha dado al tutor información recogida diligentemente y con rigor. El *feedback* ha de adecuarse a las concepciones ingenuas de aprendizaje que puede que se manifiesten en el trabajo de los estudiantes.

Concepciones del conocimiento que tienen los estudiantes

El «plan de desarrollo intelectual y ético» de Perry describe cómo van evolucionando los estudiantes a lo largo de su experiencia universitaria por lo que respecta a su comprensión de qué es el conocimiento (Perry 1970). Según él, los estudiantes empiezan pensando que hay un enorme número de respuestas correctas y que su trabajo consiste en aprenderlas y devolvérselas al profesor correctamente. Perry define este proceso de aprendizaje con la memorable frase «adición cuantitativa de corrección discreta». Describe a los estudiantes como individuos que pasan por una serie de estadios de comprensión cada vez mayor de la naturaleza del conocimiento, entre los que figura, por ejemplo, un relativismo extremo, en el cual todas las respuestas son percibidas como igualmente válidas. Un estudiante que no formula una conclusión al final de un ensayo puede estar dejándolo a criterio del lector, dado que todo tipo de conclusiones son percibidas como igualmente válidas. En ese caso, un *feedback* que simplemente diga «falta conclusión» puede que no ayude a progresar al estudiante en absoluto. El *feedback* de los profesores es a menudo (aunque no siempre) generado desde una posición epistemológica más sofisticada que la del estudiante y esto ofrece un amplio abanico de posibilidades para interpretar incorrectamente el *feedback* o para dar lugar a la más pura incompreensión.

La concepción del lenguaje de la disciplina por parte del estudiante

Lea y Street (1998) describen el caso de un estudiante que después de entregar un ensayo en una asignatura de Historia, recibió como *feedback*: «Me gustan las conclusiones de tu ensayo, es relevante y ha sido bien argumentado». Al mismo tiempo, el estudiante recibía *feedback* de otro ensayo entregado en una asignatura paralela de Antropología en el que se le criticaba tan duramente su capacidad para formular argumentos claros o concluir de forma justificada, que se le aconsejaba que recurriera a los servicios de apoyo a los estudiantes. Lea y Street interpretaban esto como consecuencia de las diferentes maneras de elaborar el discurso en cada disciplina y que en Antropología implicaba recurrir a diferentes formas de argumentación y uso de evidencias, ya que no se trataba en modo alguno de una cuestión de escritura académica inadecuada en un ensayo. Si el estudiante no dominaba el discurso de la Antropología y no lo había practicado suficientemente, entonces la ayuda para escribir ensayos en general difícilmente podía serle de utilidad, independientemente de si la ayuda provenía de un profesor o si venía de un orientador en técnicas de estudio y escritura académica.

El *feedback* ha de ser distinto en función del tipo de escritura que se espere y según lo que presumiblemente los estudiantes saben sobre él. En estructuras modulares de asignaturas es común que los estudiantes crucen las fronteras disciplinares y tengan que enfrentarse a este tipo de diferencias en el discurso. Los estudiantes de ciencias y de carreras técnicas a menudo tienen dificultades con los ensayos típicos de las ciencias sociales aunque sepan articular perfectamente un escrito en sus propias disciplinas, pero también hay profundas diferencias discursivas dentro de las ciencias sociales, por ejemplo, entre Sociología y Psicología, y dentro de las Humanidades, por ejemplo entre Historia y Literatura.

De forma parecida, Higgins *et al.* (2001) investigan los casos de comunicación fallida del *feedback*. Describen un caso en que todo el *feedback* de un tutor consistía en «Un esfuerzo suficiente. Un mayor análisis crítico de las cuestiones clave habría ayudado». El estudiante, que quería ser algo mejor que «suficiente», quedó frustrado por la pobreza del análisis crítico del tutor.

Condición 9

El *feedback* es recibido y tenido en cuenta

Hay unos cuantos estudios que relatan que los estudiantes, cuando les devuelven el trabajo corregido, echan un vistazo a la nota al final y después simplemente lo tiran a la papelera, junto con todo el *feedback*.

«Algunas veces leo de verdad los comentarios pero sé que no volveré a escribir el mismo ensayo jamás... De alguna manera tiendo a ignorarlos, a menos que haya algo realmente muy alarmante.» (Hounsell, 1987)

Crooks (1988) ha sintetizado un gran número de investigaciones sobre esta cuestión; allí donde las calificaciones de pruebas parciales o trabajos de curso contaban de manera significativa en la nota final, los estudiantes prestaban menos atención al *feedback* que las acompañaba. Jackson (1995) encontró que los estudiantes de tercer año eran particularmente propensos a mirar sólo la nota, en lugar de leer los comentarios en los ensayos. También explicaba que a los estudiantes les gusta ver el *feedback*, pero más bien para asegurarse de que se han leído su ensayo con atención y que lo han calificado justamente.

No se puede asegurar que los estudiantes leerán y prestarán atención al *feedback*, aun cuando éste haya sido formulado con todo el cariño y haya sido devuelto con la máxima celeridad. Pero sí hay una serie de pasos que han de ser seguidos para implicar a los estudiantes con el *feedback*:

- Pedir a los estudiantes que especifiquen sobre qué tipo de *feedback* quieren recibir (en relación a la actividad propuesta) y no dar *feedback* sobre nada más
- Dar el *feedback* sin la nota, de manera que los estudiantes tengan que leer el *feedback* para hacerse una idea de cómo están avanzando
- Pedir que autoevalúen sus actividades (sin que condicione la calificación) de manera que los estudiantes presten atención a si su punto de vista se corresponde con el del profesor.
- En una revisión de la bibliografía sobre autoevaluación y evaluación entre iguales, Dochy *et al.* (1999) destacan que hay evidencias en in-

investigaciones controladas de que la evaluación entre iguales mejora la actuación de los estudiantes (en comparación con un grupo de control) y que aumenta el dominio de los estudiantes de sus estrategias de aprendizaje

- Utilizar actividades en dos fases con *feedback* en la primera, con la intención de que el estudiante pueda mejorar la calidad de su trabajo entregándolo nuevamente en una segunda fase, que es la única que se califica. Cooper (2000) ha descrito cómo este sistema puede mejorar la actuación de la mayoría de los estudiantes, particularmente la de algunos de los estudiantes más flojos.
- Dar la calificación sólo después de la autoevaluación y tras haber ofrecido el *feedback* del tutor. Taras (2001) describe la efectividad de una secuencia de este tipo como componente de la evaluación sumativa.

Condición 10

El estudiante actúa ante el *feedback*

Esta cuestión tiene que ver con el impacto del *feedback* en el aprendizaje futuro. El *feedback* puede quizá corregir errores de forma precisa pero también puede seguir sin producir cambio alguno en la manera en que un estudiante realice la siguiente actividad o emprenda cualquier tarea de aprendizaje futura. Esto puede ocurrir por varios motivos:

- El *feedback* puede llegar demasiado tarde para ser puesto en práctica por los estudiantes.
- El *feedback* puede mirar sólo hacia atrás –dirigiéndose sólo a cuestiones asociadas a una materia que no volverá a ser objeto de estudio de nuevo–, en lugar de mirar hacia adelante y encarar las siguientes actividades de estudio o las tareas que el estudiante tendrá que realizar.
- El *feedback* puede ser poco realista o preciso por lo que se refiere al esfuerzo del estudiante (p. ej. «leer la bibliografía» en lugar de «para una concepción opuesta», Smith cap. 2 pág. 24-29).
- El *feedback* puede que le esté pidiendo al estudiante que haga algo que no sabe cómo hacer (p. ej. «sé más sociológico» o «exprésate con mayor claridad»).
- El *feedback* puede ser muy específico del contexto y sólo puede aplicarse a la actividad concreta en cuestión, en lugar de referirse a as-

pectos más generalizables como formas de trabajo o perspectivas que puedan ser relevantes para otras actividades.

- El *feedback* puede desanimar y conducir a esforzarse menos en los estudios en lugar de dedicarse más.
- Puede que no haya seguimiento alguno para comprobar si los estudiantes han realizado alguna acción a partir del *feedback*, de manera que se les deja que ignoren el *feedback* con total impunidad.

Ding (1998) sugiere que incluso cuando los estudiantes leen los comentarios del *feedback*, hacen muy poco con ello. Por el contrario, Brookheart (2001) encontró que los estudiantes con buenos resultados lo utilizan todo: calificaciones, *feedback* y autoevaluación, todo activamente para aprender y orientar sus actividades de estudio futuras. Las variables más importantes aquí son quizá, como tantas veces, las que tienen que ver con el estudiante más que con el profesor. Enseñar a los estudiantes a supervisar su propia actuación es, según el análisis teórico que realiza Sadler (1989) sobre el papel del *feedback*, el auténtico objetivo final del *feedback*.

Investigaciones realizadas sobre el impacto del uso de la «evaluación en clase» en universidades de Estados Unidos subrayan una y otra vez el impacto no en el aprendizaje de un contenido específico, sino en el desarrollo metacognitivo de los estudiantes y en su capacidad de incrementar el control sobre su propio aprendizaje (véase Steadman, 1998 para un resumen). Parece necesario enseñar a los estudiantes a usar el *feedback* para desarrollar el control metacognitivo (Sadler 1989). Sin embargo, la mejora de la capacidad de aprendizaje puede no tener los efectos esperados. Ramsden *et al.* (1987), estudiando el impacto de un programa de «técnicas de estudio» diseñado para conseguir que un mayor número de estudiantes adoptaran una forma de aprendizaje en profundidad, encontraron que en realidad se conseguía lo contrario. ¡Una mayor consciencia de cómo aprendían permitió a los estudiantes adoptar una forma de aprendizaje superficial en mayor medida, ya que se adaptaban a las demandas de bajo nivel que percibían en las formas de evaluación de las asignaturas que cursaban! De nuevo, esto ilustra cómo la manera de percibir la evaluación por parte de los estudiantes influye en su aprendizaje.

CONCLUSIÓN

Las condiciones bajo las que la evaluación favorece el aprendizaje están siendo verificadas en la práctica en el contexto de un proyecto a gran escala que se ha iniciado con un estudio de la evaluación aplicada en asignaturas de ciencias de dos universidades. Los profesores de las distintas asignaturas, en las que se utiliza una amplia variedad de formas de evaluación, están recogiendo datos sobre, por ejemplo, cómo distribuyen su esfuerzo los estudiantes en relación a lo que se les pide para la evaluación y cómo responden al *feedback*. Están utilizando esta información para diagnosticar problemas potenciales en sus asignaturas, introducir cambios en la evaluación para resolver estos problemas y después evaluar si los cambios han afectado positivamente la manera de aprender de los estudiantes. Es muy similar a cualquier proceso de investigación-acción que implica la mejora de la docencia, pero con una diferencia fundamental: la atención principal no está en la metodología de enseñanza, sino en la evaluación. La hipótesis de partida es que hay un margen más amplio para mejorar la docencia cambiando aspectos de la evaluación que cambiando cualquier otra cosa y, al mismo tiempo, que los profesores saben menos sobre cómo los estudiantes responden a la evaluación que sobre cualquier otro aspecto. Según vaya desarrollándose el proyecto, la comprensión de los profesores y las evidencias sobre los cambios más efectivos en las asignaturas permitirán desarrollar más estas «condiciones». Se trata de una empresa colaborativa a gran escala sobre «el conocimiento de la evaluación» que proporcionará estudios de caso sobre cambios que han resultado efectivos y, también, un marco conceptual elaborado que ayudará a explicar por qué han sido efectivos. La intención es que estas condiciones puedan ser utilizadas como una lista de control por cualquier profesor que quiera revisar la efectividad del sistema de evaluación de sus propias asignaturas en términos de favorecer el aprendizaje.

AGRADECIMIENTOS

La revisión bibliográfica que conforma la base de este artículo ha sido realizada por el Student Support Research Group de la Open University. El proyecto de investigación para el cual se ha realizado la revisión bibliográfica, «Mejora de la efectividad de la evaluación formativa en ciencias», ha sido posible gracias a la financiación de £250.000 recibida del Higher Education Funding Council for England dentro del Programa para el desarrollo de la docencia y el aprendizaje.

BIBLIOGRAFÍA

- ANGELO, T.A. y CROSS, K.P. (1993). *Classroom Assessment Techniques: a handbook for college teachers*. San Francisco, California: Jossey-Bass.
- BAIRD, L.L. (1985). «Do grades and tests predict adult accomplishment?». *Research in Higher Education*, 23 (1), 3-85.
- BERLINER, D.C. (1984). «The half-full glass: a review of research on teaching», en P.L. Hosford (ed.) *Using What We Know About Teaching*. Alexandria, Va.: Association for Supervision and Curriculum Development.
- BIGGS, J.B. y COLLIS K.F. (1982). *Evaluating the Quality of Learning: the SOLO taxonomy*. Nueva York: Academic Press.
- BLACK, P. y WILIAM, D. (1998). «Assessment and classroom learning». *Assessment in Education*, 5 (1), 7-74.
- BRIDGES, P., COOPER, A., EVANSON, P., HAINES, C., JENKINS, D., SCURRY, D., WOOLF, H. y YORKE, M. (2002). «Coursework marks high, examination marks low: discuss». *Assessment and Evaluation in Higher Education*, 27 (1), 36-48.
- BRITTINGHAM, B.E. (1988). «Undergraduate students' use of time: a classroom investigation». *To Improve the Academy*, 7, 45-52.
- BROOKHEART, S. M. (2001). «Successful students' formative and summative uses of assessment information». *Assessment and Evaluation in Higher Education*, 8 (2), 154-169.
- BRUNER, J.S. (1974). *Toward a Theory of Instruction*. Cambridge, Mass: Harvard University Press.
- CARLESS, D.M. (2002). «The 'mini-viva' as a tool to enhance assessment for learning». *Assessment and Evaluation in Higher Education*, 27 (4), 353-363.
- CARROLL, M. (1995). «Formative assessment workshops: feedback sessions for large classes». *Biomedical Education*, 23 (2), 65-67.
- CHANSARKAR, B.A. y RAUT-ROY, U. (1987). «Student performance under different assessment situations». *Assessment and Evaluation in Higher Education*, 12 (2), 115-122.
- CHICKERING, A.W. y GAMSON, Z.F. (1987). *Seven Principles to Good Practice in Undergraduate Education*. Racine, Wi.: The Johnson Foundation Inc.
- CHICKERING, A.W. y GAMSON, Z.F. (1991). *Applying the Seven Principles to Good Practice in Undergraduate Education*. San Francisco: Jossey-Bass.
- COLE, S., COATES, M. y LENTELL, H. (1986). «Towards good teaching by correspondence». *Open Learning*, 1 (1), 16-22.

- CONWAY, M.A., COHEN, G. y STANHOPE, N. (1992). «Why is it that university grades do not predict very long-term retention?» *Journal of Experimental Psychology: General*, 121 (3), 49-57.
- COOK, A. (2001). «Assessing the use of flexible assessment». *Assessment and Evaluation in Higher Education*, 26 (6), 539-549.
- COOPER, N.J. (2000). «Facilitating learning from formative feedback in level 3 assessment». *Assessment and Evaluation in Higher Education*, 25 (3), 279-291.
- CROOKS, T.J. (1988). «The impact of classroom evaluation practices on students». *Review of Educational Research*, 58 (4), 438-481.
- CURTIS, S. y SHAMI, N. (2002). «The effect of taking paid employment during termtime on students' academic studies». *Journal of Further and Higher Education*, 26 (2), 129-138.
- DING, L. (1998). *Revisiting assessment and learning: implications of students' perspectives on assessment feedback*. Ponencia presentada en el Congreso Annual de la Scottish Educational Research Association, Universidad de Dundee, 25-26 septiembre.
- DOCHY, F., SEGERS, M. y SLUIJSMANS, D. (1999). «The use of self-, peer- and coassessment: a review». *Studies in Higher Education*, 24 (3), 331-350.
- DUNKIN, M. J. (1986). «Research on Teaching in Higher Education», en M.C. Wittrock (ed.). *Handbook of Research on Teaching*, Nueva York: Macmillan. 3ª ed.
- FORBES, D. y SPENCE, J. (1991). «An experiment in assessment for a large class», en R. Smith (ed.) *Innovations in Engineering Education*. Londres: Ellis Horwood.
- GAGNE, R. M. (1977). *The Conditions of Learning*. Nueva York: Holt, Rinehart and Winston. 3ª ed.
- GARDINER, L.F. (1997). «Redesigning higher education: producing dramatic gains in student learning». *ASHE-ERIC Higher Education Reports*, 23 (7), Washington, DC: Association for the Study of Higher Education.
- GIBBS, G. (1992). *Assessing More Students*. Oxford: Oxford Centre for Staff Development.
- GIBBS, G. (1995). *Assessing Student Centred Courses*. Oxford: Oxford Centre for Staff Development.
- GIBBS, G. (1999). «Using assessment strategically to change the way students learn», en S. Brown y A. Glasner (eds.) *Assessment Matters in Higher Education*. Buckingham: Society for Research into Higher Education and Open University Press.

- GIBBS, G. y LUCAS, L. (1987). «Coursework assessment, class size and student performance: 1984-94». *Journal of Further and Higher Education*, 21 (2), 183-192.
- GIBBS, G. y SIMPSON, C. (2002). *Evaluation of Regional Retention Activity: interim report*. Open University: Student Support Research Group Report 40/2002.
- GREER, L. (2001). «Does changing the method of assessment of a module improve the performance of a student?» *Assessment and Evaluation in Higher Education*, 26 (2), 128-138.
- HATTIE, J.A. (1987). «Identifying the salient facets of a model of student learning: a synthesis of meta-analyses». *International Journal of Educational Research*, 11, 187-212.
- HIGGINS, R., HARTLEY, P. y SKELTON, A. (2001). «Getting the message across: the problem of communicating assessment feedback». *Teaching in Higher Education*, 6 (2), 269-274.
- HOUNSELL, D. (1987). «Essay writing and the quality of feedback», en J. T. E. Richardson, M.W. Eysenck y D. Warren-Piper (eds.) *Student Learning: research in education and cognitive psychology*. Milton Keynes: Open University Press and Society for Research into Higher Education.
- HUTCHINGS, P., MARCHESE, T., y WRIGHT, B. (1991). *Using Assessment to Strengthen General Education*. Washington, DC: American Association for Higher Education.
- HYLAND, F. (2001). «Providing effective support: investigating feedback to distance language learners». *Open Learning*, 16 (3), 231-247.
- INNIS, K. (1996). *Diary Survey: how undergraduate full-time students spend their time*. Leeds: Leeds Metropolitan University.
- JACKSON, M. (1995). *Making the grade: the formative evaluation of essays* [online]. UtiliBASE: disponible en <http://ultibase.rmit.edu.au/Articles/dec96/jacks1.htm>
- JAMES, D. y FLEMING, S. (2004-05). «Agreement in Student Performance in Assessment». *Learning and Teaching in Higher Education*, 1, 32-50.
- KEMBER, D., NG, S., TSE, H., WONG, E.T.T. y POMFRET, M. (1996). «An examination of the interrelationships between workload, study time, learning approaches and academic outcomes». *Studies in Higher Education*, 21 (3), 347-358.
- KNIVETON, B. H. (1996). «Student Perceptions of Assessment Methods». *Assessment & Evaluation in Higher Education*, 21 (3), 229-238.

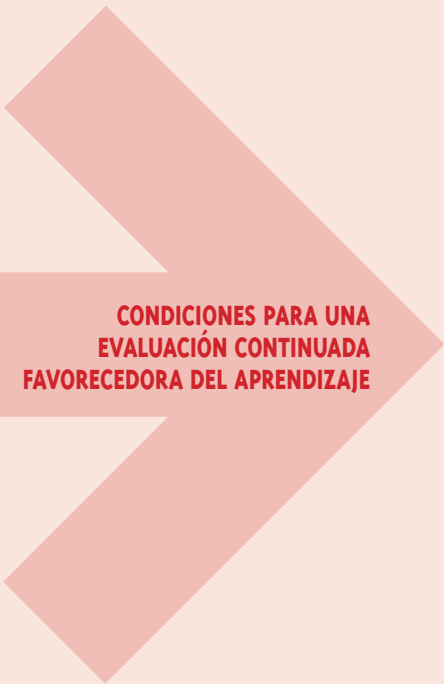
- KULIK, C.C., KULIK, J.A. y COHEN, P.A. (1980). «Instructional technology and college teaching». *Teaching Psychology*, 7, 199-205.
- LEA, M. y STREET, B. (1998). «Student writing in higher education: an academic literacies approach». *Studies in Higher Education*, 23(2), 157-172.
- MACDONALD, J. (2002). «'Getting it together and being put on the spot': synopsis, motivation and examination». *Studies in Higher Education*, 27 (3), 329-338.
- MACLELLEN, E. (2001). «Assessment for learning: the different perceptions of tutors and students». *Assessment and Evaluation in Higher Education*, 26(4), 307-318.
- MACFARLANE, B. (1992). «The 'Thatcherite' generation of university degree results». *Journal of Further and Higher Education*, 16, 60-70.
- MARTON, F. y WENESTAM, C.G. (1978). «Qualitative differences in the understanding and retention of the main points in some texts based on the principle-example structure», en M.M. Gruneberg, P.E. Moris y R.N. Sykes (eds.) *Practical Aspects of Memory*, Londres: Academic Press.
- MCKEACHIE, W.J., PINTRICH, P.R., LIN, Y. y SMITH, D. (1986). *Teaching and Learning in the College Classroom: a review of the research literature*. Ann Arbor: National Centre for Research to Improve Postsecondary Teaching and Learning, University of Michigan.
- MELTON, R. (2000). *Assessment and Assignments — a review of related issues based on data collected from the 1998 courses survey*, Report SRC 200, Open University.
- MENTKOWSKI, M., ASTIN, A., EWELL, P. y MORAN, E. T. (1991). *Catching Theory up with Practice: conceptual frameworks for assessment*. Washington DC: AAHE Assessment Forum.
- MILLER, C.M.I. y PARLETT, M. (1974). *Up to the Mark: a study of the examination game*. Guildford: Society for Research into Higher Education.
- MOFFAT, M. (1989). *Coming of Age in New Jersey: college and American culture*. New Brunswick, NJ: Rutgers University Press.
- ORSMOND, P., MERRY, S. y REILING, K. (2002). «The use of exemplars and formative feedback when using student derived marking criteria in peer and self assessment». *Assessment and Evaluation in Higher Education*, 27 (4), 309-323.
- PATON-SALTZBERG, R. y LINDSAY, R. (1993). *The Effects of Paid Employment on the Academic Performance of Full-time Students in Higher Education*. Oxford: Oxford Polytechnic.

- PENNY, A.J. y GROVER, C. (1996). «An analysis of student grade expectations and marker consistency». *Assessment and Evaluation in Higher Education*, 21 (2), 173-184.
- PERRY, W.G. (1970). *Forms of Intellectual and Ethical Development in the College Years*. Nueva York: Holy, Rinehart and Winston.
- RAMSDEN, P. (1991). «A performance indicator of teaching quality in higher education: the Course Experience Questionnaire». *Studies in Higher Education*, 16, 129-150.
- RAMSDEN, P. (1992). *Learning to teach in Higher Education*. Londres: Routledge.
- RAMSDEN, P., BESWICK, D. y BOWDEN, J. (1987). «Learning processes and learning skills», en J.T.E. Richardson, M.W. Eysenck y D.W. Warren-Piper (eds.) *Student Learning: research in education and cognitive psychology*. Milton Keynes: Open University Press and Society for Research into Higher Education.
- RICE, M., MOUSLEY, J. y DAVIS, R. (1994). «Improving student feedback in distance education: a research report», en T. Evans y D. Murphy (eds.) *Research in Distance Education 3: revised papers from the third research in distance education conference*. Geelong, Victoria: Deakin University Press.
- ROBERTS, D. (1996). «Feedback on assignments». *Distance Education*, 17 (1), 95-116.
- ROWNTREE, D. (1987). *Assessing Students — how shall we know them?* Londres: Kogan Page.
- SADLER, D.R. (1989). «Formative assessment and the design of instructional systems». *Instructional Science*, 18, 119-144.
- SADLER, D.R. (1998). «Formative assessment: revisiting the territory». *Assessment in Education*, 5 (1), 77-84.
- SÄLJÖ, R. (1982). *Learning and Understanding*. Goteborg: Acta Universitatis Gothoburgensis.
- SAMBELL, K. y MCDOWELL, L. (1998). «The construction of the hidden curriculum: messages and meanings in the assessment of student learning». *Assessment and Evaluation in Higher Education*, 23 (4), 391-402.
- SCHUNK, D. (1984). «Self-efficacy perspective on achievement behaviour». *Educational Psychologist*, 19, 48-58.
- SCHUNK, D. (1985). «Self-efficacy and classroom learning». *Psychology in the Schools*, 22, 208-223.
- SCOULER, K. (1998). «The influence of assessment method on students' learning approaches: multiple choice question examinations vs. essay assignment». *Higher Education*, 35, 453-472.

- SCOULER, K. y PROSSER, M. (1994). «Students' experiences of studying for multiple choice question examinations». *Studies in Higher Education*, 19 (3), 267-280.
- SNYDER, B.R. (1971). *The Hidden Curriculum*. Cambridge, MA: MIT Press.
- SLY, L. (1999). «Practice tests as formative assessment improve student performance on computer managed learning assessments». *Assessment and Evaluation in Higher Education*, 24 (3), 339-344.
- STARR, J.W. (1970). «Student opinion on methods of assessment». *Educational Review*, 22, 243-253.
- STEADMAN, M. (1998). «Using classroom assessment to change both teaching and learning». *New Directions for Teaching and Learning*, 75, 23-35.
- STEPHENSON, K., SANDER, P. y NAYLOR, P. (1996). «Student perceptions of the tutor's role in distance learning». *Open Learning*, 11 (1), 22-30.
- TANG, C. (1994). «Effects of modes of assessment on students' preparation strategies», en G. Gibbs (ed.) *Improving Student Learning: theory and practice*. Oxford: Oxford Centre for Staff Development, 151-170.
- TARAS, M. (2001). «The use of tutor feedback and student self-assessment in summative assessment: towards transparency for students and for tutors». *Assessment and Evaluation in Higher Education*, 26 (6), 605-614.
- THOMAS, J.W., IVENTOSH, L. y ROHWER, W.D. (1987). «Relationships among student characteristics, study activities and achievement as a function of course characteristics». *Contemporary Educational Psychology*, 12, 344-364.
- TINTO, V. (1993). *Leaving College: rethinking the causes and cures of student attrition*. Chicago: University of Chicago Press. 2ª ed.
- TORRANCE, H. y PRYOR, J. (1998). *Investigative Formative Assessment: teaching, learning and assessment in the classroom*. Buckingham: Open University Press.
- TYNJALA, P. (1998). «Traditional studying for examination vs constructivist learning tasks: do learning outcomes differ?» *Studies in Higher Education*, 23 (2), 173-191.
- VOS, P. (1991). *Curriculum Control of Learning Processes in Higher Education*, 13th International Forum on Higher Education of the European Association for Institutional Research, Edinburgh.
- WARREN, J.R. (1971). *College Grading Practices: an overview, Report No. 9*. Washington D.C., ERIC Clearinghouse on Higher Education.
- WOOTTON, S. (2002). «Encouraging learning or measuring failure?» *Teaching in Higher Education*, 7 (3), 353-357.

WOTJAS, O. (1998). «Feedback? No, just give us the answers». *Times Higher Education Supplement*, 25 septiembre 1998.

YORKE, M. (2001). «Formative assessment and its relevance to retention». *Higher Education Research and Development*, 20 (2), 115-126.



**CONDICIONES PARA UNA
EVALUACIÓN CONTINUADA
FAVORECEDORA DEL APRENDIZAJE**

Graham Gibbs ha sido director del Oxford Learning Institut de la Universidad de Oxford hasta 2007, al que sigue vinculado como investigador visitante, ha dirigido el Student Support Research Group de la Open University. Anteriormente codirigió el Centro para la Práctica en Enseñanza Superior de la Open University y fue profesor y director académico del Centro de Desarrollo y Aprendizaje en Oxford Brookes. Desde 1997 ha trabajado en el desarrollo e implementación de estrategias de docencia y aprendizaje en el Equipo de Coordinación Nacional para el Desarrollo de la Enseñanza y del Aprendizaje de la HEFCE (Higher Education Funding Council for England). Autor de numerosas publicaciones sobre docencia, aprendizaje y evaluación en educación superior, es miembro cofundador del Improving Student Learning Symposia y del International Consortium for Educational Development in Higher Education (ICED) y ha desarrollado una importante actividad como asesor en más de 120 instituciones de educación superior en todo el mundo. Ha recibido diversas distinciones como doctor honoris causa, la más reciente por la Universidad de Utrecht por su contribución al desarrollo de la docencia y el aprendizaje en educación superior.

Claire Simpson es Máster en Métodos de búsqueda psicológica; investigadora del Grupo de búsqueda en apoyo al estudiante (Student Support Research Group) en la Open University (2001); responsable de los servicios de evaluaciones de la Business School de la Open University (2003), y asesora sobre métodos de búsqueda, diseño de cuestionarios, selección de muestras, análisis y procesos de datos para los proyectos relacionados con el apoyo al estudiante.

Ha colaborado en la evaluación del proyecto de retención de estudiantes, búsqueda sobre tipos de *feedback* del docente e impacto en el aprendizaje del estudiante. También realiza formación relativa a la búsqueda educativa.