

2

EVALUACIÓN FORMATIVA DE LOS APRENDIZAJES EN EL CONTEXTO UNIVERSITARIO: RESISTENCIAS Y PARADOJAS DEL PROFESORADO

(FORMATIVE ASSESSMENT IN HIGHER EDUCATION: TEACHERS' RESISTANCE AND PARADOX)

Leonor Margalef García
Universidad Alcalá de Henares

DOI: 10.5944/educxx1.17.2.11478

Cómo referenciar este artículo/How to reference this article:

Margalef García, L. (2014). Evaluación formativa de los aprendizajes en el contexto universitario: Resistencias y paradojas del profesorado. *Educación XXI*, 17 (2), 35-55. doi: 10.5944/educxx1.17.2.11478

Margalef García, L. (2014). Formative Assessment in Higher Education: Teachers' resistance and paradox. *Educación XXI*, 17 (2), 35-55. doi:10.5944/educxx1.17.2.11478

RESUMEN

En este artículo se presentan los resultados de uno de los objetivos de un proyecto de investigación sobre evaluación formativa: analizar las resistencias y paradojas a las que se enfrenta el profesorado universitario para implementar procesos de evaluación orientados a mejorar el aprendizaje de sus estudiantes. La investigación se fundamentó en un estudio de casos múltiples conformado por profesorado de las diferentes ramas de conocimiento de la Universidad de Alcalá.

Nuestros hallazgos sugieren que estas resistencias y paradojas tienen que ver con las creencias y concepciones del profesorado acerca de la enseñanza, el aprendizaje y la evaluación, la influencia de la cultura organizativa y las condiciones institucionales; a lo que hay que añadir la inseguridad e incertidumbre que provocan los procesos de innovación en las prácticas evaluadoras. En el estudio se concluye que esta incertidumbre no solo se debe a una falta de información y comprensión de los procesos de cambio evaluativos sino, también, a que el profesorado cuenta con escasos referentes y aprendizajes experienciales en evaluación formativa en los que apoyarse.

PALABRAS CLAVE

Evaluación de los aprendizajes, evaluación formativa, innovación, educación superior.

ABSTRACT

This article presents the results of one of the goals of a research project on formative assessment: to analyze the resistances and paradoxes faced by the faculty members in order to implement processes of learning-oriented assessment of their students. We present a multiple case study on teachers from different knowledge areas of Alcalá University.

Our findings suggest that these resistances and paradoxes have a close relationship with teachers' beliefs and conceptions about teaching, learning and assessment, the influence of organizational culture and institutional conditions and additionally the insecurity and uncertainty caused by innovation in the processes of evaluative practices.

The study concludes that this uncertainty is not only due to a lack of information and understanding of the change in evaluation but also due to the scarce references and experiences in formative assessment on which to rely on.

KEY WORDS

College Outcomes Assessment, Performance Based Assessment, Educational Innovation, Higher education.

INTRODUCCIÓN

Los estudios e investigaciones educativas evidencian que los sistemas de evaluación que emplea el profesorado son decisivos en el modo en el que el alumnado aprende/estudia y tienen una gran influencia en el contenido que se enseña y en cómo se enseña (Álvarez Méndez, 2001, 2008; Biggs, 2006; Black *et al.* 2003; Brown y Glasner, 2003; Gimeno, 2008; Margalef, 2005, 2009).

Los resultados que se presentan en este artículo son fruto de una investigación¹ más amplia, centrada en dos grandes ejes: la enseñanza por competencias y la evaluación formativa. El propósito principal fue reflexionar sobre el sistema actual de evaluación de los aprendizajes para conocer las concepciones y creencias, comprender la perspectiva de los formadores e identificar algunos de los factores que contribuyen o dificultan el cambio en las creencias, los métodos y procedimientos de evaluación. El cambio que se propone en el modo de entender la enseñanza y el aprendizaje en la universidad conlleva necesariamente una transformación de la concepción de la evaluación del aprendizaje, de sus finalidades y funciones. Sin embargo,

en la universidad, la evaluación sigue ejerciendo una función predominante de control, selección y clasificación de los estudiantes.

En este artículo solo se abordan los hallazgos relacionados con el segundo eje de la investigación, el de la evaluación formativa, y con uno de los objetivos del proyecto: analizar las resistencias y paradojas del profesorado para involucrarse en procesos de evaluación formativa y auténtica del aprendizaje de sus estudiantes.

¿DESDE DÓNDE PARTIMOS?

Esta investigación se sustentó en la concepción de la evaluación para el aprendizaje o evaluación formativa. Wiliam (2011) realiza una exhaustiva revisión de los estudios sobre evaluación formativa y evaluación para el aprendizaje, concluyendo que no hay claridad conceptual ni consenso en el modo de definir ambos conceptos. Esta es una de las causas que dificulta la comprensión de los efectos que este tipo de evaluación genera en las prácticas docentes o en la mejora del aprendizaje de los estudiantes. Este autor concluye que la evaluación contribuye al aprendizaje cuando tiene lugar durante el proceso de enseñar y aprender; es realizada por profesores, pares y estudiantes que buscan reflexionar, tomar decisiones y responder a la información y evidencias recogidas, con la finalidad de mejorar la enseñanza y el aprendizaje y comprometer a los estudiantes en el desarrollo de sus aprendizajes futuros.

Coincidimos con estas características esenciales pero, además, asumimos la evaluación como una actividad crítica de aprendizaje, que no distingue entre actividades de aprendizaje y actividades de evaluación (Álvarez Méndez, 2001, 2008); integrada con tareas auténticas (Monereo, 2009) orientada al aprendizaje autónomo y autorregulado (Carles 2006, 2007); una evaluación para la comprensión (Brown y Glasner, 2003; Gibbs y Simpson, 2009; Taras, 2008), que contribuye a la emancipación de los estudiantes, a su desarrollo personal, moral, académico y social (Black *et al*, 2003; Boud y Falchikov, 2006; Margalef, 2005, 2009). Es decir, una evaluación que se caracteriza por ser holística, democrática, ética e ideológica.

Desde un modelo de enseñanza-aprendizaje basado en la construcción del conocimiento y del aprendizaje activo y reflexivo del alumnado se deben ampliar los ámbitos y dimensiones de la evaluación. En ese sentido, se extiende no solo al dominio de conceptos, principios o leyes, sino a las habilidades intelectuales relacionadas con la comprensión, la aplicación y el razonamiento (comparación, relación de ideas, argumentación, proposición, contrastación, la reconstrucción del sentido y el significado a partir de

la información). También integra la evaluación de habilidades complejas de resolución de problemas nuevos y la creación y producción de conocimiento (Biggs, 2006; Brown y Glasner, 2003; Stobart, 2010). Asimismo, es necesario incluir la evaluación de habilidades sociales relacionadas con el trabajo en equipo, la colaboración y la empatía. Entra en juego, además, la evaluación de actitudes como el respeto hacia la posición de los otros, la escucha activa, la tolerancia, el compromiso en la tarea y la responsabilidad con el grupo (Boud y Falchikov, 2006).

Tal y como apunta el estudio de Kember, Leung y Ma (2007), hay que crear un ambiente de aprendizaje que integre tanto el aprendizaje como la evaluación, en y fuera de la clase y tener en cuenta los factores contextuales (sociales, psicológicos y pedagógicos) que influyen en el desarrollo del aprendizaje.

Pero si nos preguntamos ¿qué hace formativa a la evaluación?, comprobamos que en la revisión de los estudios hay coincidencias en señalar que es el *feedback*, el juicio crítico, la participación activa del alumnado y el diálogo sobre lo aprendido (Carless, 2007; Gibbs y Simpson, 2009; Stobart, 2010). Handley y Williams (2011) señalan en las conclusiones de su investigación que para mejorar el *feedback* y lograr un efecto positivo sobre el aprendizaje es esencial facilitar el diálogo entre profesores (tutores) y estudiantes y trabajar en los procesos de interacción. Asimismo, otros autores consideran las dimensiones relacionadas con las características y condiciones de los criterios de evaluación: claridad, transparencia, construcción conjunta, apropiación por parte de los estudiantes (Mutch, 2003; Weaver, 2006). En este sentido, algunos estudios exploran el potencial de las rúbricas para orientar el *feedback* y favorecer la autoevaluación (Jonsson and Svingby, 2007); mientras otros se centran en el tipo de *feedback* que se genera y en los efectos y respuestas que provocan en la mejora del aprendizaje de los estudiantes (Shute, 2008).

Con este marco teórico intentamos indagar hasta qué punto este modo de concebir la evaluación estaba presente en la concepción y en las prácticas del profesorado universitario que formaba parte de los estudios de casos de nuestro proyecto y cuáles eran las resistencias y las paradojas para poder avanzar en su puesta en práctica.

CONTEXTO DEL ESTUDIO: METODOLOGÍA Y PROCESO DE LA INVESTIGACIÓN

Para favorecer la comprensión de los resultados presentados en este artículo explicamos brevemente el proceso de investigación y los métodos

de recogida de datos. Las preguntas de investigación giraron en torno a las siguientes cuestiones: ¿qué modelos/sistemas de evaluación predominan en las diferentes ramas de conocimiento?, ¿en qué concepciones se sustentan?, ¿qué competencias se evalúan?, ¿qué condicionamientos externos e internos facilitan u obstaculizan los sistemas de evaluación integrados y alternativos?

La investigación se fundamentó en un estudio de casos múltiple (Stake, 2000; 2006; Yin, 2006). Estos casos estuvieron conformados por grupos de profesores y profesoras de la Universidad de Alcalá que procedían de distintos contextos y áreas de conocimiento, pero que compartían un compromiso activo con la innovación, así como una sensibilidad por la búsqueda y construcción de procedimientos y estrategias alternativas de evaluación.

a) Caso 1: Experto en Docencia Universitaria

Conformado por 23 profesores matriculados en el Experto en Docencia Universitaria (2007/2008): 10 de Ciencias e Ingenierías, 6 de Ciencias Sociales y Jurídicas, 5 de Ciencias de la Salud y 2 de Humanidades. En cuanto a la categoría docente: 32% profesores ayudantes, 18% profesores titulares de universidad, 27% profesores contratados doctor y colaboradores y el resto, profesorado asociado y personal investigador.

b) Caso 2: Grupos de Innovación Docente

Profesorado de distintas áreas de conocimiento adscritos a grupos de innovación, involucrados en experiencias y proyectos de innovación de sus propias prácticas docentes. En total, se contactó con los 202 profesores de los 26 grupos a los que se les remitió el cuestionario. De ellos, 9 grupos eran de Ciencias Sociales y Jurídicas, 5 de Ciencias, 4 de Ingenierías, 3 de Ciencias de la Salud y 5 de Humanidades. En cuanto a la categoría docente el 50% son Profesores Titulares; el 16% profesores ayudantes doctor; el 8% catedráticos y el resto pertenecen a otras categorías de contratados y asociados.

c) Caso 3: Encuentro de Innovación

Profesorado y alumnado participantes del III Encuentro de Innovación en Docencia Universitaria dedicado exclusivamente al debate y la reflexión en torno a los nuevos sistemas de evaluación en la enseñanza universitaria. Sesenta participantes, entre profesorado y estudiantes, pertenecientes a las ramas de Ciencias Sociales y Jurídicas (24), Ciencias de la Salud (12),

Ciencias (10), Ingenierías (8) y de Humanidades (6), formaron parte de los grupos de discusión.

Los datos: recogida y análisis

Los datos se recogieron a través de diferentes técnicas e instrumentos, como se detalla en la tabla 1. Se utilizó una combinación de técnicas cualitativas y cuantitativas de acuerdo a los objetivos y preguntas de investigación. El propósito de esta diversidad fue facilitar la triangulación de los resultados obtenidos a través del estudio de los distintos casos.

Técnica empleada	Cantidad	Descripción general
<i>Análisis de Documentos (Experto en Docencia Universitaria)</i>	23 programas 23 fichas de autoanálisis	Revisión y análisis de las propuestas de evaluación de los programas y guías docentes. Autoanálisis de concepciones de evaluación, elaborado por el profesorado participante.
<i>Talleres (Experto en Docencia Universitaria)</i>	1 taller (23 profesores)	Presentación de las características de la evaluación formativa. Posibilidades de aplicación de estrategias de evaluación formativa.
<i>Grupos de Discusión (Experto en Docencia Universitaria)</i>	3 grupos (entre 6 y 8 participantes por grupo)	Debate acerca de las concepciones y prácticas de evaluación del profesorado de distintas ramas de conocimiento en base al autoanálisis realizado por cada uno de los participantes.
<i>Grupos de Discusión (Encuentro de Innovación)</i>	4 grupos (60 seleccionadas entre los 140 participantes)	Análisis DAFO de los sistemas de evaluación formativa realizados por profesorado y estudiantes.
<i>Cuestionario</i>	52 cuestionarios	Cuestionario en el que se recogen las concepciones y prácticas del profesorado relacionadas con la evaluación por competencias y las necesidades de formación para implementar sistemas alternativos de evaluación en la universidad.

Tabla 1: Técnicas de recogida de datos

El análisis de los datos en una primera fase se llevó a cabo de manera diferenciada. Posteriormente, se procedió a integrar toda la informa-

ción a través de un proceso de triangulación. El análisis cuantitativo de los datos del cuestionario se realizó a través de las frecuencias de respuesta de acuerdo a las categorías establecidas previamente. El análisis cualitativo se llevó a cabo con el soporte del software ATALAS-ti para facilitar el proceso de codificación y organización de las categorías. El procedimiento de categorización siguió un proceso inductivo y deductivo. En el primer caso, se crearon nuevas categorías emergentes que surgían del «texto» para formar parte del sistema de categorías. El proceso deductivo partió de la utilización de categorías previamente construidas. Una vez completado el análisis, se construyó un mapa conceptual empleando las principales categorías para dar sentido a los datos y contar con un primer modelo explicativo de la interrelación entre categorías. Para el análisis se tuvo en cuenta el muestreo teórico y las comparaciones constantes de modo que se analizaban hasta que se producía una saturación de los datos. También para asegurar la confiabilidad y credibilidad se apeló a la triangulación de fuentes, métodos y datos.

PRESENTACIÓN DE RESULTADOS Y DISCUSIÓN

En este artículo presentamos sólo la dimensión de las resistencias y paradojas, fruto de la triangulación de los casos antes expuestos, principalmente del análisis e interpretación de esta categoría en los grupos de discusión, las actividades y diarios del profesorado, así como de las respuestas abiertas del cuestionario². En el gráfico 1 se sintetizan las categorías y subcategorías que se desarrollarán en los siguientes apartados.

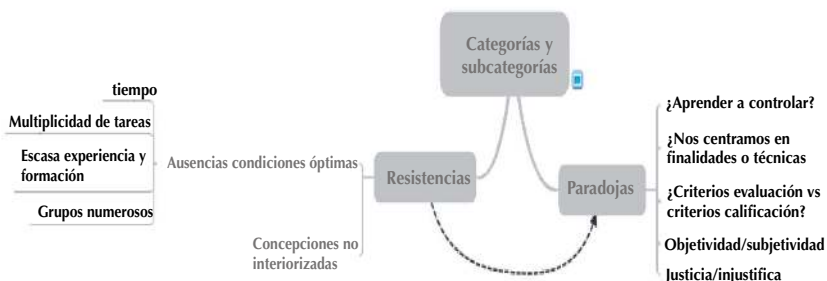


Gráfico 1: Categorías de análisis

Resistencias ante la evaluación formativa

Algunas de las resistencias que identificamos se podrían atribuir a las lógicas «resistencias al cambio», mientras que otras están justificadas por lo difícil que resulta innovar los procedimientos de evaluación teniendo en cuenta toda una serie de factores externos que impiden que las condiciones de aplicación sean las óptimas. Igualmente, encontramos resistencias concretas al ampliarse el sentido de la evaluación más allá de los contenidos conceptuales, o al considerar el proceso y diversificarse las estrategias e instrumentos. Muchas veces, surgen al amparo de una noble justificación de «ser pragmáticos y realistas con el sistema universitario actual».

La siguiente cita ejemplifica este sentido de resistencia al cambio, motivado por la inseguridad que dicho cambio genera tanto al profesorado como al alumnado:

Profesora: Yo veo una cierta resistencia. No todos... aparte de que nosotros tenemos que cambiar la actitud, la responsabilidad y que tendremos que colaborar todos... también en muchos alumnos hay una cierta resistencia igual que en nosotros, o sea, lo conocido queda igual, y lo nuevo crea miedo. Hay cierta resistencia porque es nuevo, crea más trabajo y da más inseguridad (GDE)³.

Aunque se plantean asuntos importantes que considerar cuando se lleva a cabo una innovación, se enfatiza la inseguridad que genera y la conciencia de la propia resistencia a la hora de experimentar dicho miedo ante lo desconocido.

Resistencia ante la ausencia de condiciones óptimas

Justificando esa resistencia general al cambio, es frecuente que se argumente no intentar nuevos procedimientos de evaluación aludiendo a la falta de condiciones óptimas para llevarlas a cabo. De acuerdo con el profesorado y alumnado que participó en la investigación, estas condiciones están relacionadas con:

- *El tiempo que demanda una evaluación de naturaleza formativa, tanto para el profesorado como para el alumnado. En el caso del profesorado exige una mayor dedicación para hacer un seguimiento continuo de los procesos de aprendizaje del alumnado: exige mayor dedicación a la tarea docente ya que obliga a ser sistemático y metódico durante el curso (GDI, P7). En el caso del alumnado, la dificultad de comprometerse en un sistema de evaluación formativa se relaciona con la presencia de un currículum sobrecargado que provoca una multiplicidad de actividades de aprendizaje y evalua-*

ción: en algunos momentos es tal la saturación de trabajos y actividades que coinciden en la misma fecha que por más que quieras no puedes dedicarle el tiempo necesario ni hacerlos con la calidad que se merecen o realmente te gustaría (GDI, A8).

Esta dificultad se acrecienta aún más debido al carácter comprimido de las asignaturas y a la existencia de asignaturas compartidas por varios profesores en las que, sin una estrecha coordinación, es imposible llevar a cabo una evaluación de proceso: *no hay cultura de coordinación con los otros profesores por lo que no se sabe lo que cada uno exige y puede generar sobrecarga en los alumnos y pérdida de calidad en los trabajos (GDI, P5).*

- *La necesidad de compaginar tareas de docencia e investigación que, junto al escaso reconocimiento de la tarea docente y al desconocimiento de la docencia en la valoración del profesorado, llevan a que estos se vean ante la encrucijada de tener que distribuir su tiempo u optar entre las tareas de enseñanza y sus actividades de investigación. En consecuencia, la docencia y la iniciativa del profesorado por introducir cualquier tipo de innovación en su práctica queda librada a la «responsabilidad y compromiso» de cada uno.*
- *La falta de experiencia y formación en procedimientos de evaluación formativa. El profesorado no se siente suficientemente preparado para afrontar un nuevo sistema de evaluación. Una profesora expresa: nos sentimos perdidos, no sabemos cómo evaluar por competencias, no tenemos información sobre procedimientos, existen pocas fuentes bibliográficas, las experiencias son aún incipientes y no tenemos disponibles suficientes resultados (GDI, P3).*
- *Elevado número de alumnos en clases. Un excesivo número de alumnos suele ser la razón más expresada por el profesorado a la hora de justificar la imposibilidad de aplicar nuevas prácticas de evaluación. La mayoría parece estar de acuerdo en que es necesario realizar otro tipo de evaluación, más individualizada, que valorara el desempeño, que promoviera la comprensión, el desarrollo del pensamiento crítico, etc., en lugar de reducirse a evaluar productos para poner notas. No obstante, dado el elevado número de alumnos, el profesorado considera que «evaluar el producto» y poner notas es lo único que se puede hacer: una cosa que compartimos, es que nos gustaría evaluar un poquito más cierta capacidad de enfrentarse a cosas nuevas, de resolver casos o problemas pero que no sean del tipo habitual. Problema: que tenemos muchos alumnos en clase y que esta actividad, para que funcione, hay que tutorizarla bien y no estamos acostumbrados a ello (GDE, P12).*

Aunque las condiciones que se describen son importantes y es preciso buscar estrategias para mejorarlas, en vez de generar alternativas que superen los obstáculos aludidos partiendo de la realidad que se vive en el aula, las condiciones se convierten en una excusa razonable para resistir la innovación, para seguir evaluando como siempre se ha hecho.

Lo que hay que hacer o la resistencia ante concepciones no interiorizadas

Un aspecto frecuente que refleja también las resistencias a la hora de aplicar innovaciones en la evaluación son las frases que expresan lo que se hace frente a lo que se tendría o no se tendría que hacer. Esta manera de expresarse con la fórmula «tengo que» evidencia una concepción que todavía no se ha interiorizado, se percibe como algo externo que «hay que hacer» casi como una obligación. En sí, refleja la dicotomía entre las concepciones interiorizadas y aquellas que todavía no se han interiorizado y deberían seguirse. Refleja la tensión entre la realidad y el ideal, un ideal que no se termina de asumir porque, entre otras cosas, existe cierta resistencia para asumirlo, como se refleja en la siguiente cita:

Profesor/a 1: Para nosotros sigue predominando el examen final en nuestra rama de conocimiento, pero creemos que empieza a cobrar un poco más de fuerza, a día de hoy, la evaluación continua, que requiere mucho más tiempo y esfuerzo por parte de los docentes y que depende de la materia y del número de alumnos.

Profesor/a 2: Una cosa es el ideal y otra cosa lo que realmente es. (GDE)

En definitiva, se observa la tensión entre la realidad y el ideal. Sin duda, no basta con asumir nuevas teorías curriculares, sino hay que transferirlas al conocimiento práctico y lograr que se transformen en acción; de lo contrario, en las prácticas cotidianas seguimos manteniendo el *statu quo*.

Muchas de estas resistencias expresan la naturaleza dilemática de los procesos educativos y enfrentan al profesorado a una serie de tensiones y paradojas.

Paradojas

En este estudio, detectamos claramente las paradojas que surgen entre las dos concepciones de evaluación: una, formativa, centrada en la calidad de la experiencia vivida y en el proceso de aprendizaje, y otra, acreditativa, con fuerte peso en la especificación de resultados y niveles de dominio o

estándares de calidad del producto (Taras, 2008). La paradoja que vive con mayor intensidad el profesorado participante de esta investigación es la dificultad de conciliar una evaluación cualitativa con énfasis en el seguimiento y en la calidad del aprendizaje, con una retroalimentación cualitativa como fuente de diálogo e información, y una evaluación cuya función es la calificación, con un fuerte peso en el control, la medición de resultados y los estándares de comparación.

En ese sentido, se amplían las conclusiones que se han obtenido en otras investigaciones sobre las percepciones de profesores universitarios, que muestran que «los profesores no tienen claridad en su comprensión sobre la evaluación formativa, sumativa y la autoevaluación, ni son consistentes cuando establecen las relaciones entre ellas. La gran contradicción existe en la definición de cómo relacionan la evaluación formativa y la sumativa en el proceso de evaluación» (Taras, 2008, p. 188).

Para no correr el riesgo de caer en un reduccionismo simplificador de procesos complejos, es coherente exponer los resultados reflejando estas tensiones y paradojas.

¿Evaluar para aprender o controlar?

Sin duda alguna, no es lo mismo evaluar que examinar (Álvarez Méndez, 2001). Examinar va asociado a realizar un examen, pero este es un recurso, un procedimiento y, como tal, podemos prescindir de él y buscar otros modos o procedimientos. Lo que está claro es que no podemos evaluar si no hay aprendizaje, pues ¿qué estaríamos evaluando? A esto se añade otra cuestión importante para entender esta paradoja: ¿qué entendemos por aprender? Dependerá de cómo concibamos los procesos de enseñanza-aprendizaje para identificar cómo entendemos la evaluación. En lo que hay consenso es en la imposibilidad de disociar la enseñanza, el aprendizaje y la evaluación.

Ahora bien, hemos visto que en el ámbito educativo todo es «evaluable», lo que no significa que sea numéricamente evaluable o cuantificable. Como indica Bain (2007), está comprobado que las recompensas externas producen efectos efímeros y pérdida de interés. Además, si nos acostumbramos a que los estudiantes se interesen solo por la nota, no contribuiremos a un aprendizaje duradero y fomentaremos —como los denomina Bain— «aprendices estratégicos», entendidos como aquellos estudiantes que desarrollan habilidades para obtener buenas calificaciones pero evitan los desafíos y el desarrollo de una comprensión más profunda; temen cometer errores y descartan buscar otras soluciones o alternativas, lo que reduce su

creatividad. Así lo expresaba un alumno en el Grupo de Discusión: *la evaluación final o tradicional no facilita el aprendizaje, no genera interés, y cuando hay poca motivación e implicación lo que te interesa es aprobar* (GDI A3).

Justamente, las competencias complejas que consideramos que son las que merecen la pena desarrollar en la enseñanza universitaria resultan mucho más difíciles de cuantificar y requieren de otros procedimientos y criterios de evaluación. Criterios que son más divergentes y ambiguos, a la vez que menos sujetos a puntajes fragmentados, y que requieren procesos y juicios holísticos. Ello provoca una tensión en el profesorado cuya concepción sigue perpetuando perspectivas técnicas que no facilitan procesos de evaluación auténtica. Generalmente, su modo de entender la evaluación es como control, y sus actos de evaluación se reducen a la sumatoria de notas y calificaciones que dejan poco margen a la reconstrucción de conocimientos, a la autoevaluación y al fomento de la autonomía por parte del alumnado. La búsqueda de evidencias está regularmente basada en criterios de medición. Cuesta aún asumir que podemos buscar información y evidencias sobre el proceso, los sujetos, lo que aprenden y cómo lo aprenden.

La concepción de evaluación formativa entendida como proceso que forma al alumnado pero también al profesorado, que se convierte en aprendizaje al permitir que se cometan errores y se aprendan de ellos y que se exploren e indaguen posibles alternativas para un desarrollo cognitivo, aún es incipiente. Desde una «lectura crítica (hermenéutica), evaluar es conocer, es contrastar, es dialogar, es indagar, es argumentar, es deliberar, es razonar» (Álvarez Méndez, 2001, p. 59). Pero vemos que esta lectura crítica aún no está instaurada suficientemente en las prácticas educativas que hemos indagado. Muchas de ellas se inclinan hacia el control y responden a otros intereses que no se adecúan a la evaluación formativa.

Las concepciones más empleadas por el profesorado que participó en esta investigación se basan, principalmente, en concebir la evaluación como un proceso dirigido a medir los resultados de los estudiantes. Como vimos, esta medición puede realizarse para cumplir varios propósitos: (1) verificar y controlar qué han aprendido y/o memorizado (y proceder a calificar); (2) retroalimentar y optimizar su proceso de aprendizaje así como comprobar la metodología de enseñanza que se está siguiendo. Sin embargo, encontramos cada vez más la presencia de profesores y profesoras que transitan hacia la evaluación como aprendizaje, tal y como lo plantean Álvarez Méndez (2008), Carless (2007), Gibbs y Simpson (2009) y Taras (2008). Este profesorado va resolviendo muchas paradojas y desafíos planteados por sus creencias y concepciones, que, en muchos casos, están fuertemente arraigadas. En estos casos, la tensión se presenta en el terreno de las prácticas

evaluativas. Esto podemos verlo en la siguiente reflexión de un profesor en el taller de evaluación:

...Hasta ahora la evaluación del trabajo la realizábamos los profesores, pero creo que sería muy importante contar con una evaluación grupal por parte de cada uno de los integrantes de los diferentes grupos para que valoraran cómo ha sido el proceso de aprendizaje del grupo y creo que una metodología muy práctica sería una que ayude a reflexionar sobre cómo se ha desarrollado el proceso, de dónde partían y a dónde han llegado. Esto nos quedaba sin evaluar, y sólo nos fijábamos en el resultado final (trabajo bueno, malo...) pero no cómo y por qué el grupo había llegado a obtener ese resultado (TE, P12).

¿En qué nos centramos: las finalidades de la evaluación o las técnicas de evaluación?

El problema de la evaluación no es un problema de técnicas, sino del porqué y para qué evaluamos. Es en este sentido cuando la evaluación se convierte en un *efecto mariposa* (Gimeno, 2008); no se limita solo a técnicas ni se ocupa del cómo evaluar, sino que va más allá y desencadena una serie de reflexiones y consecuencias prácticas. Pensar en las finalidades nos lleva a replantearnos el para qué evaluamos y qué funciones cumple en mi enseñanza. Y es aquí donde comienzan a replantearse algunas funciones, lo que provoca que se interfieran en las otras y desencadena una reflexión más profunda y global.

En el profesorado participante de la investigación, se ha detectado una gran preocupación por los instrumentos, por ejemplo, el profesorado dedica mucha atención a la planificación del examen, a la búsqueda de sofisticados modos de puntuar o a la utilización de sistemas de calificaciones basados en ecuaciones, combinaciones matemáticas o parámetros excluyentes. En general, el debate o la preocupación por innovar en la evaluación comienzan por los procedimientos y el refugio en que estas técnicas puedan brindar seguridad, objetividad y rigor. No obstante, poco a poco se comprende que lo esencial es tener claros los propósitos y finalidades, el para qué evalúo. Cuando el profesorado asume este punto de partida, comienza a comprender la importancia de pensar sobre qué, para qué, por qué evaluar y con posterioridad puede diseñar los procedimientos adecuados para ello. En este sentido, apreciamos que, es en este momento, cuando se integran las diversas dimensiones y contenidos a evaluar en un mismo procedimiento, se desarrolla una visión más holística del proceso y se supera la fragmentación de múltiples tareas.

Stobart (2010) refuerza estas ideas cuando indica que lo fundamental es insistir en el aprendizaje fundado en principios. En este contexto, la adecuación a la finalidad se comprobaría examinando si la evaluación aborda no solo los contenidos sino las finalidades de la asignatura. Por ello, la consecuencia lógica de este modo de analizar la coherencia entre finalidades, principios y propuestas de evaluación es que necesitamos un amplio repertorio de estrategias de evaluación, y aquí es donde en nuestro estudio hemos detectado una gran debilidad en las prácticas de evaluación del profesorado. En los resultados del cuestionario, se detectó que en las ramas de ciencias, ciencias de la salud y arquitectura e ingeniería se suelen utilizar con mayor frecuencia los exámenes (tipo test o a desarrollar), ejercicios de prácticas de laboratorios o cuadernos de prácticas; mientras que en las ciencias sociales se comparten con otras estrategias como trabajos individuales o grupales y presentaciones de los estudiantes en clase. Sin embargo, en todas las ramas de conocimiento se otorga un mayor peso o distribución en los criterios de calificación a los resultados de los exámenes, siendo en algunos casos prácticamente simbólica la valoración del conocimiento a través de otras estrategias que manifiestan utilizar y considerar en sus prácticas evaluativas.

Criterios de evaluación versus criterios de calificación

Otra tensión muy común es la que se da entre los criterios de calificación y los criterios de evaluación. En la mayoría de los casos hemos apreciado cómo estos se intercambian muy fácilmente; tanto es así que bajo el epígrafe *criterios de evaluación* se listan los criterios de calificación, especificando generalmente los porcentajes de calificación para cada una de las dimensiones o para los instrumentos de evaluación. En diversas ocasiones, no se especifican criterios sino procedimientos o instrumentos de evaluación con un peso determinado en la calificación (por ejemplo el 50% para el examen final, el 20% para el trabajo y así sucesivamente). Es interesante detectar que los criterios de evaluación no forman parte del saber práctico del profesorado y que les resulta una dimensión que no tienen incorporada a su práctica docente. Cuando se refieren a las competencias que podrían ser la base para los criterios de evaluación, estos no se especifican en coherencia con las competencias propuestas. Un profesor lo expresa como un gran dilema de su práctica: «*La mayor dificultad la tengo a la hora de definir los criterios de evaluación de forma coherente con los objetivos de la materia, con el diseño de las tareas que me permitan evaluar esos criterios y con su expresión en criterios de calificación coherente con los mismos*» (ACE, P1)

Hemos comprobado que existe una gran dificultad para evaluar competencias de distinta naturaleza. El peso mayor en los criterios lo tienen las competencias relacionadas con la adquisición de conocimientos, la asi-

milación de los mismos y la aplicación, aunque esta suele restringirse a la resolución de problemas en contextos más reproductivos que creativos. En algunos casos, se ha detectado que se infravaloran competencias a las que se ha dedicado atención, tiempo y trabajo durante el proceso de enseñanza-aprendizaje. Este salto tiene que ver con la ambigüedad y la incertidumbre que producen las competencias y actividades complejas, flexibles y abiertas.

¿Hacer frente a las inseguridades: objetividad/subjetividad, justicia-injusticia?

La evaluación formativa desborda al profesorado no solo por falta de preparación, formación o referentes experienciales sino por las concepciones y creencias predominantes sobre la evaluación, en definitiva, sobre cómo conciben el conocimiento. El profesorado, a través de sus intervenciones en los diferentes casos, ha dejado claro que ha sido formado y socializado en la búsqueda de objetividad, en la comprobación de saberes y en la evaluación como calificación. Esto mismo lo destaca Martín al mencionar el resultado de algunas investigaciones que revelan los supuestos más arraigados en las concepciones del profesorado sobre el conocimiento, el aprendizaje y la evaluación: «...los docentes seguimos interpretando en muchas ocasiones el aprendizaje como una apropiación directa de la realidad, una copia del objeto de conocimiento... bajo el principio de certeza: el ser humano puede acceder a la verdad y los estudiantes adquirir el conocimiento correcto» (2009, p. 203). Por tanto, la concepción de evaluación coherente con estas premisas es entenderla como un medio objetivo para comprobar esa apropiación del conocimiento, lo que el estudiante ha adquirido.

Las formas de entender la evaluación como actividad crítica de aprendizaje, implica asumir otra concepción de conocimiento y aprendizaje, caso contrario se genera una fuerte resistencia ante la pérdida de control y ante el miedo a no ser estrictamente objetivos. Mucho más ante la posibilidad de compartir la responsabilidad de la evaluación con el alumnado.

Como indica Shaw en el fondo subyacen controversias profundas, sospechas, miedos, inseguridades que se asientan profundamente en la búsqueda de mediciones, de la norma, de la selección y clasificación: «Es una orientación hacia resultados tangibles de la práctica que lleva incluso a convertir lo cualitativo en cuantitativo para dar mayor rigor y objetividad» (2003, p. 21). Así lo expresan los propios profesores:

Mi principal inquietud en el momento de la evaluación es que esta sea justa y en este sentido me preocupan, de manera particular, tanto la idoneidad del método de evaluación y su consonancia efectiva con los

objetivos que persigue la asignatura, como el buen uso que yo mismo sepa hacer de él, en lo que concierne, por ejemplo, al diseño de las pruebas de evaluación, a su nivel de exigencia y una calificación objetiva y equitativa (ACE).

...ser «justo» a la hora de calificar a nuestros alumnos supongo que es una preocupación generalizada. Lo de la ecuanimidad creo que lo he abandonado por imposible... me resulta difícil, si no imposible, ser ecuaníme (ACE).

Esto trae también aparejado la pérdida de los factores situacionales que se asocian con subjetividad y la búsqueda del respaldo de evaluaciones estandarizadas. Como advierte Stobart «se corre el riesgo de interpretar lo estandarizado como justo (o sea igual para todos) sin investigar si lo que se evalúa es o no injusto para algunas de las personas que se someten a la prueba» (2010, p. 206).

También hemos comprobado cómo se produce la asociación de algunos conceptos e ideas. Entre las más arraigadas cabe citar la asociación de subjetividad con arbitrariedad y la de objetividad con justicia o equidad. Recordemos que una creencia generalizada que suele apoyar los procedimientos tradicionales de evaluación es considerar los procedimientos subjetivos como arbitrarios y, por extensión, injustos. Pero lo cierto es que en algunos grupos de discusión el profesorado ha llegado a comprender que, como afirma Álvarez Méndez (2008), se puede ser «objetivamente injusto». En muchos casos se ha puesto de relieve el temor del profesorado a no ser objetivos, por tanto, a no ser justos.

CONCLUSIÓN: ¿QUÉ RIESGOS AFRONTAR?

Uno de los grandes retos para asumir una práctica evaluadora como actividad crítica de aprendizaje es hacer frente a la inseguridad que pone de manifiesto el profesorado para asumir propuestas evaluativas más complejas, menos conocidas y poco experimentadas por ellos mismos. En esta investigación se ha detectado que no hay formación previa en métodos y procedimientos más divergentes y abiertos, lo que genera incertidumbre en el profesorado. Principalmente, la mayor inseguridad es dejar el escudo de la objetividad que brindan las pruebas con respuestas convergentes y asumir que el conocimiento es complejo, inestable, dependiente del contexto de aplicación y que, además, puede resultar de una nueva situación o provocar la creación de nuevos conocimientos. El profesorado ha manifestado que se siente más seguro al refugiarse y sentirse respaldado por normas, prácticas habituales y costumbres, ya que sienten que las innovaciones en el terreno

de la evaluación generan resistencias no solo porque es nuevo, sino porque no tienen experiencias o referentes en lo que apoyarse. Hemos visto que en diversas ocasiones las resistencias se originan por una falta de información y comprensión de los procesos de cambio. Al no aparecer con claridad las concepciones epistemológicas que suponen un cambio en la naturaleza, producción y distribución del conocimiento, al no contar con referentes que se puedan debatir, contrastar, adaptar o aprender, ni orientaciones y condiciones que faciliten estos procesos se genera más resistencia entre el profesorado, que no comparte los principios de la evaluación formativa.

En nuestra investigación hemos comprobado que existe un condicionante Institucional, a nivel macro, como señalan Dysthe y Engelsen (2011); nos referimos a las reglamentaciones internas (normativas de evaluación) y externas (por ejemplo, reales decretos, el estatuto del estudiante). En el caso de nuestra universidad, la nueva normativa es muy general y deja autonomía a los centros y al profesorado, establece finalidades y apoya procesos de evaluación continua y formativa. Pero cada uno la lee o la interpreta desde sus propios marcos referenciales y se afana en encontrar el modo legitimado para evaluar, así como un respaldo y seguridad ante cualquier conflicto que pudiera surgir. La evaluación formativa permanece, en gran medida, en un terreno que despierta sospecha de «subjetividad» y este es un gran desafío para el profesorado que la pone en práctica. «El individualismo, la orientación de la evaluación en relación a la norma es todavía ampliamente dominante. A pesar de ello hay algún cambio, y este no opera a través de la política sino a través de la experiencia vivida por el profesorado y el alumnado» (Boud y Falchikov, 2006, p. 411).

Es importante, por lo tanto, actuar no solo en el plano de la política curricular, organizativa y estructural, sino también en proporcionar experiencias vividas en nuevas formas de evaluar para ir logrando cambios en la esfera más profunda. Una vía que hemos encontrado es mediante el fomento de acciones formativas y apoyo a la innovación en el campo evaluativo. Acciones basadas en el aprendizaje experiencial, a través de las cuáles podemos acceder a las creencias (teorías implícitas), y al conocimiento explícito del profesorado para contribuir mediante teorías alternativas, psicopedagógicas e interdisciplinarias a reflexionar sobre su práctica, reinterpretarla y lograr la reconstrucción de ese conocimiento práctico. Este aprendizaje relevante de nuestro profesorado es esencial si aspiramos a una transformación real de las prácticas evaluativas que contribuya a una mejora en el aprendizaje de nuestros estudiantes.

Notas

- ¹ Este proyecto ha sido llevado a cabo por el Grupo de Investigación FIT (Formar, Indagar, Transformar) de la Universidad de Alcalá y financiado en el marco de la convocatoria de Ayudas a la Creación y Consolidación de Grupos de Investigación (CCG07/UAH/HUM-2075) de la Comunidad de Madrid y la Universidad de Alcalá.
- ² El cuestionario tenía 26 preguntas, de las cuales cuatro eran abiertas. Se estructuró en los siguientes bloques: concepción sobre competencias, prácticas de evaluación por competencias, concepción de evaluación, prácticas de evaluación, implementación de nuevos sistemas de evaluación, dificultades para implementar una evaluación alternativa y necesidades formativas. En este artículo, solo se hace referencia a los resultados de las preguntas abiertas de los tres últimos bloques y que se incluyen en las categorías que aparecen en el gráfico 1 de resistencias y paradojas.
- ³ Códigos: GDE = Grupo de Discusión – Experto Docencia Universitaria. GDI, P = Grupo de Discusión Encuentro Innovación – Profesor (a). GDI, A = Grupo de Discusión Encuentro Innovación – Alumno(a). TE = Taller Experto Docencia Universitaria. ACE = Autodiagnóstico concepciones de evaluación. PA = Programa de la asignatura.

REFERENCIAS BIBLIOGRÁFICAS

- Álvarez Méndez, J. M. (2001). *Evaluar para conocer, examinar para excluir*. Madrid: Morata.
- Álvarez Méndez, J. M. (2008). Evaluar el aprendizaje en una enseñanza centrada en competencias. En J. Gimeno (comp.) *Educación por competencias, ¿qué hay de nuevo?* (pp. 206-234). Madrid: Morata.
- Bain, K. (2007). *Lo que hacen los mejores profesores universitarios*. Valencia: Servicio de Publicaciones Universidad de Valencia.
- Biggs, J. (2006). *Calidad del aprendizaje universitario*. Madrid: Narcea.
- Black, P.; Harrison, C.; Lee, C.; Marshall, B. y William, D. (2003). *Assessment for Learning: Putting it into practice*. Maidenhead: Open University Press.
- Boud, D. y Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31(4), 399-413.
- Brown, S. y Glasner, A. (2003). *Evaluar en la universidad. Problemas y nuevos enfoques*. Madrid: Narcea.
- Carless, D. et al (2006). Learning-oriented assessment: principles and practice. *Assessment & Evaluation in Higher Education*, 31(4), 395-398.
- Carless, D. (2007). Learning-oriented assessment: conceptual bases and practical implications. *Innovations in Education and Teaching International*, 44(1), 57-66.
- Dysthe, O. y Engelsen, K. S. (2011). Portfolio practices in higher education in Norway in an international perspective: macro-, meso- and micro-level influences. *Assessment & Evaluation in Higher Education*, 36(1), 63-79.
- Gibbs, G. y Simpson, C. (2009). *Condiciones para una evaluación continuada favorecedora del aprendizaje*. Barcelona: Octaedro.
- Gimeno Sacristán, J. (comp.) (2008). *Educación por competencias, ¿qué hay de nuevo?* Madrid: Morata.
- Handley, K. y Williams, L. (2011). From copying to learning: using exemplars to engage students with assessment criteria and feedback. *Assessment & Evaluation in Higher Education*, 36(1), 95-108.
- Jonsson, A. y Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2, 130-144.
- Kember, D; Leung, D. y Ma, R. (2007). Characterizing learning environments capable of nurturing generic capabilities in Higher Education. *Research in Higher Education*, 48(5) 609-632.
- Margalef, L. (2005). Los retos de la evaluación auténtica en la enseñanza universitaria: coherencia epistemológica y metodológica. *Perspectiva Educativa*, 45, 25-44.
- Margalef, L. (2009). Aprender a evaluar: transformar nuestras prácticas universitarias. En M. Castelló (ed) *La evaluación auténtica en la enseñanza secundaria y universitaria: Investigación e Innovación*. (pp. 87-115). Barcelona: Edebé.
- Martín, E. (2009). Profesorado competente para formar alumnado competente: El reto del cambio docente. En J. Pozo y M. P. Pérez (coords) *Psicología del aprendizaje universitario: La formación en competencias*. (pp. 199-215). Madrid: Morata.

- Monereo, C. (2009). La autenticidad de la evaluación. En M. Castelló (ed) *La evaluación auténtica en la enseñanza secundaria y universitaria: Investigación e Innovación*. (pp. 15-32). Barcelona: Edebé.
- Mutch, A. (2003). Exploring the practice of feedback to students. *Active Learning in Higher Education*, 4(1), 24-28.
- Shaw, I. (2003). *La evaluación cualitativa*. Barcelona: Paidós.
- Shute, V. (2008). Focus on Formative Feedback. *Review of Educational Research*, 78(1), 153-189.
- Stake, R. (2000). Cases Studies. En N. Denzin y Y. Lincoln (Eds.) *Handbook Qualitative Research*. (pp. 435-454). London: Sage.
- Stake, R. (2006). *Evaluación comprensiva y evaluación basada en estándares*. Barcelona: Graó.
- Stobart, G. (2010). *Tiempo de pruebas: Los usos y abusos de la evaluación*. Madrid: Morata.
- Taras, M. (2008). Formative and summative assessment: Perceptions and Realities. *Active Learning in Higher Educations*, 9(2), 172-192.
- Weaver, M. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education*, 31(3), 379-394.
- William, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37, 3-14.
- Yin, R.K. (2006). Case study methods. En J.L. Green, G. Camilli y P.B. Elmore (Eds.) *Handbook of complementary methods in education research*. (pp. 111-122). Washington, D.C: American Educational Research Association.

PERFIL ACADÉMICO Y PROFESIONAL DE LA AUTORA

Leonor Margalef García, Doctora en Ciencias de la Educación, es profesora titular del Departamento de Didáctica de la Universidad de Alcalá. Es Vicerrectora de Calidad e Innovación Docente y ha estado a cargo de la Formación del Profesorado Universitario. Es coordinadora del Grupo de Investigación: Formar, Indagar, Transformar. Sus líneas de investigación están centradas en los procesos de innovación curricular en Educación Superior; la formación y el desarrollo profesional del profesorado y la evaluación en la enseñanza universitaria.

Dirección postal de la autora: Facultad de Educación
Aulario María Díaz de Guzmán
C/. San Cirilo s/n
Alcalá de Henares - 28071 - Madrid
E-mail: Leonor.margalef@uah.es

Fecha Recepción del Artículo: 16. Febrero. 2012

Fecha Modificación del Artículo: 14. Abril. 2012

Fecha Aceptación del Artículo: 15. Abril. 2012

Fecha de Revisión para publicación: 09. Enero. 2014