

# Nuevas Vías de Aprendizaje y Enseñanza y sus Consecuencias: una Nueva Era de Evaluación<sup>1</sup>

Filip Dochy, Mien Segers y Sabine Dierick  
Universidad de Lovaina, Bélgica  
Universidad de Maastricht, Países Bajos

<sup>1</sup> Traducción: Raúl Alelú Paz.  
Revisión: Carmen Vizcarro Guarch.  
Universidad Autónoma de Madrid.

## Abstract

In the present contribution the author outlines the changes that are being considered in learning and instruction, and their effects on student assessment. Globally, these effects lead to what might be considered a burst in the assessment culture. He argues that new needs in education force us to think about new assessment criteria and that educational innovation is not possible without dramatic changes in assessment.

Keywords: learning assessment, teaching and assessment innovative methods, problem based learning, portfolios.

## Resumen

En el presente artículo vamos a esbozar los cambios que actualmente se consideran claves en aprendizaje e instrucción, y sus efectos en la evaluación de los estudiantes. El total de estos efectos puede ser considerado como el comienzo de la cultura de la evaluación. Se defenderá que las nuevas formas de concebir la educación exigen nuevos criterios de evaluación y la innovación educativa no es posible sin plantear importantes cambios en la evaluación.

Palabras clave: evaluación del aprendizaje, métodos innovadores de enseñanza y evaluación, aprendizaje basado en problemas, portafolios.

### 1. Desarrollos recientes en la sociedad del aprendizaje

Algunos colegas estarán sin duda en desacuerdo cuando afirmo que la sociedad del aprendizaje cambiará más en la década que viene de lo que ha cambiado en las pasadas décadas. Trataré de convencer con mis palabras a alguno de ellos. Los recientes desarrollos en la sociedad del aprendizaje están determinados por al menos seis desarrollos diferentes (Dochy y McDowell, 1997).

La edad de la información, introduciéndonos en el siglo XXI, está caracterizada por una cantidad de información infinita, dinámica y cambiante. La información se intercambia rápidamente y el conocimiento crece

de manera exponencial. No hay posibilidad de que un científico abarque todo el conocimiento que se genera dentro de su disciplina. Los científicos, actualmente y en el futuro, necesitan llegar a dominar el conocimiento básico de su campo y las habilidades necesarias para navegar por sus disciplinas. Estos cambios son también ciertos para los profesores, que tradicionalmente han sido reconocidos como las fuentes de todo el conocimiento y experiencia dentro del proceso educativo. El profesor ahora se representa más bien como una llave para abrir la puerta de dominio del conocimiento y la experiencia.

Al entrar en una nueva era de posibilidades tecnológicas, la educación hará uso la autopista electrónica en todas sus formas. Las vías electrónicas de comunicación se emplean ya de forma intensiva en el contexto académico. El uso de los multimedia, redes locales, sistemas de comunicación compartidos, Internet, bases de datos electrónicas compartidas, facilidades de videoconferencia, materiales de autoaprendizaje electrónicos, estudio guiado y apoyado por redes, sistemas de evaluación continua, sistemas de ingreso y seguimiento, etc, nos conducirán al desarrollo de nuevas estrategias de enseñanza y aprendizaje.

El uso de la tecnología de la información hace que el mundo sea más pequeño. También contribuyen a ello los modernos medios de transporte. Ciertamente, el nuevo mercado común europeo ya establecido, el Euro como la moneda única europea y la continua expansión de la Unión Europea hacia los países del Este contribuirán a una mayor movilidad tanto de las personas como de la información.

Recientemente, se ha experimentado un rápido cambio en el mercado laboral en el que se demanda una mano de obra más flexible y en el que aumentan los trabajos por períodos cortos, a tiempo parcial e informales. Existe un incremento de la presión por parte de la industria para que la educación superior proporcione graduados que sean inmediatamente empleables y efectivos en los negocios y en la industria (Moerkerke, 1996). Las organizaciones de empleo han señalado un acuerdo insuficiente entre los resultados de los programas de estudio y las necesidades del mercado de trabajo (Harvey & Green, 1994). Si se requiere que los sistemas de educación vocacional y académica formen licenciados que inmediatamente puedan ser utilizados en el ámbito laboral, ello implica una necesidad creciente de procedimientos para evaluar adecuadamente las habilidades relevantes (De Rijke & Dochy, 1995). Los especialistas en evaluación educativa (Birenbaum, 1996; Shavelson, Xiaohong & Baxter, 1996) han reconocido que, en un futuro próximo, se tendrán que desarrollar sistemas efectivos y eficientes para realizar evaluaciones.

La ideología y las estrategias del mercado libre se han introducido recientemente en el mundo de la educación. Cada vez se están tomando más iniciativas por parte de los gobiernos para fomentar una educación dirigida a la demanda, tratando de equilibrar las demandas de la industria y las preferencias de los estudiantes. Los centros de educación superior se ven abocados cada vez más a competir unos con otros para conseguir estudiantes, y como resultado de ello las necesidades y deseos de los estudiantes reciben una mayor atención, situando de esta manera a los estudiantes en el papel de consumidores (Green, Brannigan, Mazelan & Giles, 1994).

Una mayor necesidad de aprender durante toda la vida en las sociedades modernas (Sambell & Mc Dowell, 1997) aumentará también la necesidad de aprendizaje a lo largo de toda la vida profesional (Dochy, Segers & Sluijsmans, 1999). En la última década, aproximadamente, se ha producido un fuerte incremento en los esfuerzos de formación en la industria y los negocios. Las presiones económicas que requerían una reestructuración del mercado de trabajo fue uno de los factores principales que llevaron al gobierno y a los empresarios a enfatizar la importancia de la adaptabilidad dentro del mercado de trabajo y la capacidad de los empleados para adquirir nuevas habilidades a lo largo de toda su vida laboral. Es un hecho ampliamente aceptado que la necesidad de un aprendizaje a lo largo de toda la vida aumentará aún más rápidamente en un futuro próximo (Moerkerke, 1996).

## 2. Cambios en los objetivos de la educación académica

Durante muchos años, el objetivo principal de la educación académica ha sido formar estudiantes con muchos conocimientos dentro de un determinado dominio. La tarea principal consistía en construir un almacén de

conocimientos básicos. Desarrollos recientes en la sociedad actual han cambiado de forma notable estos objetivos. Se hace énfasis en la formación de individuos con un alto nivel de conocimientos, pero también con habilidades para resolver problemas, habilidades profesionales y aprendizaje auténtico, por ejemplo, el aprendizaje en contextos de la vida real:

"... El funcionamiento satisfactorio en este área requiere personas autónomas, adaptables y pensantes, que aprenden de forma autónoma, capaz de comunicarse y cooperar con otros. Las competencias específicas que se requieren de estas personas incluyen:

- a. competencias cognitivas tales como la resolución de problemas, pensamiento crítico, formulación de preguntas, búsqueda de información relevante, elaboración de juicios informados, uso eficiente de la información, realizar observaciones e investigaciones precisas, inventar y crear cosas nuevas, analizar datos, habilidades comunicativas para presentar los datos, expresión oral y escrita.
- b. competencias metacognitivas tales como autorreflexión y autoevaluación
- c. competencias sociales tales como dirigir discusiones y conversaciones, persuasión, cooperación, trabajo en grupos, etc, y
- d. disposiciones afectivas, por ejemplo, perseverancia, motivación interna, responsabilidad, autoeficacia, independencia, flexibilidad, afrontamiento de situaciones frustrantes (Birenbaum, 1996, p.4).

El objetivo principal de la educación superior ha cambiado para dirigirse a los estudiantes con el fin de proporcionar apoyo para desarrollarse como "practicantes reflexivos" capaces de reflexionar críticamente acerca de su propia práctica profesional (Falchikov & Boud, 1989; Kwan & Leung, 1996). Los estudiantes que ocupan posiciones en las organizaciones modernas han de ser capaces de analizar la información, mejorar su habilidades de resolución de problemas y comunicación y reflexionar sobre su propio papel en el proceso de aprendizaje. Las personas tienen que ser capaces de adquirir conocimiento de manera independiente y de emplear ese cuerpo de conocimientos organizados para resolver problemas imprevistos (Dochy y col., 1999).

### 3. La enseñanza tradicional y la cultura del examen

Dadas las demandas actuales, debemos desarrollar entornos de aprendizaje más potentes, que abarquen tanto enseñanza como evaluación. Son necesarias una enseñanza y una evaluación apropiadas (Glaser, Lesgold & Lajoie, 1987).

El enfoque tradicional de la enseñanza consideraba a los aprendices como receptores pasivos de la información. La memorización del contenido, narrado por el profesor, era el objetivo principal del proceso de enseñanza. El conocimiento almacenado sólo era abstracto. El aprendizaje y la enseñanza se consideraban procesos individuales con el profesor individual situado enfrente de un auditorio, compuesto por una conjunto de estudiantes individuales (Segers, Dochy & De Corte, 1999; Dochy & McDowell, 1997).

El enfoque de evaluación que acompañaba este enfoque de enseñanza se concentraba principalmente en la evaluación de conocimientos básicos, supuestamente adquiridos a través de experiencias tediosas y experiencias que exigían ensayos y repeticiones enseñados en clase o en el libro de texto. Durante las pasadas tres décadas, el desarrollo de pruebas con fines de rendición de cuentas, así como su corrección e interpretación estuvo dominada por expertos en evaluación que empleaban sofisticados modelos psicométricos. En el mundo occidental, especialmente en los Estados Unidos, su trabajo estaba guiado por las demandas de objetividad y equidad, exigiéndose un alto grado de estandarización debido a las importantes consecuencias asociadas con el resultado de estas pruebas. En estas circunstancias, los tests, principalmente los de elección múltiple, verdadero / falso o de

emparejamiento de elementos, fueron las herramientas típicas para la evaluación (Birenbaum, 1996).

Este sistema de evaluación se denomina en ocasiones la "cultura de los tests o del examen" (Klenasser, Horsch & Tastad, 1993; Wolf, Bixby, Glenn & Gardner, 1991), y presenta las siguientes características:

- La enseñanza y la evaluación se consideraban actividades separadas, la primera la responsabilidad del profesor y la última responsabilidad del experto en medida.
- El plan de la prueba, la redacción del elemento así como el desarrollo de criterios para evaluar la realización del test y el proceso de puntuación normalmente no son compartidos con los estudiantes y son un misterio para ellos.
- Los ítems son con frecuencia sintéticos y con frecuencia no están relacionados con la experiencia cotidiana de los estudiantes.
- Las pruebas suelen adaptarse a un formato de elección y examinan el conocimiento de unidades descontextualizadas y discretas de un determinado dominio científico.
- Los tests suelen ser de lápiz y papel, se administran en clase con fuertes limitaciones de tiempo y se prohíbe el recurso a materiales y herramientas de apoyo y ayuda.

En otras palabras, el primer borrador del trabajo de los estudiantes, producido en condiciones estresantes y limitaciones poco realistas, se usa con frecuencia para determinar consecuencias de gran alcance. Lo que se evalúa es meramente el producto, sin tener en cuenta el proceso, y el informe de los resultados normalmente presenta la forma de una única puntuación total (Birenbaum, 1996).

Estos instrumentos han recibido muchas críticas (Birenbaum, 1996; Collins, 1990; Glaser & Silver, 1994; Hambleton & Murphy, 1992; Magone, Cai, Silver & Wang, 1994; Masters & Mislevy, 1993; Wolf y col., 1991). Una crítica es que estos exámenes tradicionales no se parecen a las tareas reales de aprendizaje; otra, es que estas pruebas no parecen reflejar adecuadamente la capacidad de resolución de problemas. Una prioridad de la práctica convencional de evaluación consiste en centrarse en conocimientos que pueden cuantificarse con facilidad más que en habilidades y conocimientos complejos. Sin embargo, estos instrumentos de medida en su mayoría no parecen evaluar habilidades cognitivas de nivel superior, tales como la capacidad de resolución de problemas, el pensamiento crítico y el razonamiento. Esto se debe, en parte, a la mentalidad de "una respuesta correcta" (Magone y col., 1994, p.317). Los exámenes convencionales con frecuencia no permiten múltiples respuestas correctas mientras que la mayoría de los problemas o tareas de la vida real no tienen una única solución ya que pueden interpretarse desde perspectivas diferentes. Estas pruebas tradicionales generalmente fomentan la memorización mas que la comprensión. Además tales tests se centran en componentes aislados, discretos del dominio en cuestión y no en las relaciones entre esos componentes. Una crítica adicional se refiere a la influencia de la práctica de evaluación sobre los procesos de enseñanza. Las pruebas tradicionales tienden a restringir los procesos de aprendizaje a un proceso de mero consumo del conocimiento proporcionado por el profesor (esto es, el enfoque tradicional de la enseñanza) (Segers y col., 1999).

#### **4. El enfoque actual de la enseñanza y la cultura de la evaluación**

Desarrollos recientes en las sociedades avanzadas nos han llevado a adoptar nuevos métodos que están en consonancia con estos desarrollos.

En oposición al enfoque tradicional, la enseñanza actual y la concepción de la evaluación acentúan la importancia de la adquisición de competencias específicas de naturaleza cognitiva, metacognitiva y social (Dochy

& Moerkerke, 1997; Segers, 1999). Feltovich, Spiro y Coulson (1993) recurren al concepto de comprensión para describir el foco principal del enfoque actual de la enseñanza y la evaluación. Estos autores definen la comprensión como "la adquisición y retención de una red de conceptos y principios sobre algún dominio que representa con exactitud los fenómenos claves y sus interrelaciones, y que pueden utilizarse de forma flexible cuando resulte pertinente para llevar a cabo objetivos diversos, con frecuencia novedosos. (p.181).

Para alcanzar este objetivo de conocimientos profundos se necesitan nuevos métodos de enseñanza. Una extensa investigación en psicología cognitiva ha influenciado de forma notable el proceso de enseñanza (Segers y col. 1999). Como ha afirmado Mislevy (1996): "La revolución cognitiva es un *fait accompli* en psicología, y ha empezado a influenciar los modos en que los educadores buscan caracterizar, observar y estimular el aprendizaje de los estudiantes" (p.411). De Corte (1990) se refiere al diseño de entornos de aprendizaje potentes indicando que éstos se caracterizan por la convicción de que el aprendizaje significa construir conocimientos y habilidades sobre la base del conocimiento previo. Un entorno de aprendizaje potente se caracteriza por un buen equilibrio entre aprendizaje por descubrimiento y exploración personal por una parte, y la enseñanza sistemática y guiada por la otra. Un aspecto importante, es que el estudiante es responsable de su propio aprendizaje. El profesor es el iniciador y el guía de este proceso. Ella o él actúa como un mentor o entrenador que proporciona oportunidades a los aprendices para que usen lo que ya conocen para entender nuevos temas. Se espera que el profesor proporcionará tareas que supongan retos significativos para el estudiante, relacionadas con su experiencia y capaces de mejorar sus estrategias de aprendizaje y comprensión. (Birenbaum, 1996). Otra característica importante de los entornos de aprendizaje potentes es la necesidad de ubicar el aprendizaje en situaciones y contextos de la vida real (Segers, 1999; Dochy & McDowell, 1997). Diversos resultados de la investigación (Vygotski, 1978; Resnick, 1987) apoyan la conclusión de que las actividades de construcción de conocimiento de los estudiantes deberían estar preferiblemente situados en contextos ricos en fuentes y materiales de aprendizaje, que ofrezcan amplias oportunidades para la interacción social y que sean representativos del tipo de tareas y problemas en que los aprendices tendrán que aplicar su conocimiento y habilidades en el futuro.

Los diferentes objetivos y los nuevos métodos de instrucción así como los nuevos hallazgos e ideas sobre los entornos de aprendizaje más potentes señalan la necesidad de reconceptualizar también la evaluación y de examinar de manera crítica la teoría subyacente (Glaser, 1990, Lohman, 1993).

El aprendizaje y la evaluación necesitan ser coherentes. La investigación ha mostrado que la naturaleza de las tareas de evaluación influye en el enfoque que los estudiantes adoptan para aprender. Los enfoques de evaluación tradicionales pueden tener efectos contrarios a los deseados (Beckwith, 1991). Segers (1997) aporta dos razones adicionales que explican porqué la enseñanza y la evaluación deben estar vinculadas. La primera de ellas es que los resultados de los estudiantes proporcionan información que puede ser utilizada para mejorar la práctica educativa sólo cuando los instrumentos que miden los resultados encajan exactamente con las prácticas de enseñanza (English, 1992). Debido a su naturaleza estática y orientada al producto, los exámenes tradicionales no proporcionan información diagnóstica relevante, necesaria para adaptar la enseñanza a las necesidades del aprendiz (Campione & Brown, 1990; Dochy, 1994; Snow & Lohman, 1989). En segundo lugar, las pruebas son ayudas diagnósticas sólo cuando pueden identificar en qué medida se están consiguiendo los objetivos. Esto significa que la evaluación debe ser capaz de reflejar en qué medida los estudiantes son capaces de usar el conocimiento de manera interrelacionada al analizar y resolver problemas auténticos.

La pruebas de evaluación inadecuadamente vinculadas a la instrucción han conducido a consecuencias no deseadas tales como una información inadecuada sobre el progreso en el aprendizaje y dificultades de aprendizaje, reducción de la motivación del estudiante para aprender, y la evaluación incorrecta de la efectividad de la enseñanza.

De esta manera, la evaluación debe ir más allá de la medida de la reproducción del conocimiento (Dochy et al., 1999), ya que las pruebas tradicionales no son apropiadas para formas de aprendizaje que se refiere a la construcción de significados por parte del estudiante y el desarrollo de estrategias para abordar nuevos problemas y tareas de aprendizaje. Es necesario, por lo tanto, que la evaluación incremente el uso de diversos tipos de elementos, para cuya solución los estudiantes tengan que interpretar, analizar, evaluar problemas y explicar sus argumentos (Segers y col, 1999).

Este enfoque de evaluación se denomina en ocasiones "cultura de evaluación" (Wolf y col., 1991; Kleinasser y col., 1993) y tiene las siguientes características:

- Fuerte énfasis en la integración de la evaluación y la enseñanza. La mayoría de los especialistas en evaluación como Birenbaum (1996), Nitko (1995) y Keeves (1994) defienden que, usada apropiadamente, la evaluación educativa pueden considerarse como una herramienta que mejore el proceso de enseñanza. El punto de vista que sostiene que la evaluación de los logros de los estudiantes es algo que ocurre exclusivamente al final del proceso de aprendizaje no es sostenible (Dochy y McDowell, 1997).
- El estudiante debe ser considerado un participante activo que comparte la responsabilidad en el proceso, practica la autoevaluación, la reflexión y la colaboración, y lleva a cabo un diálogo continuo con el profesor. El estudiante participa en el desarrollo de los criterios para la evaluación de sus resultados.
- Tanto el proceso como el producto, deben ser evaluados.
- La evaluación adopta muchas formas, todas ellas generalmente consideradas por el enfoque psicométrico como "evaluaciones no estandarizadas realizadas en el curso de la enseñanza" (Koretz, Stecher, Klein & McCaffrey, 1994).
- Normalmente no hay presión de tiempo, y se permite hacer uso de diversas herramientas que se emplean en la vida real para realizar tareas semejantes.
- Las tareas son con frecuencia interesantes, significativas y auténticas, incluyendo a menudo investigaciones de varios tipos
- Los estudiantes documentan sus reflexiones en un diario y usan portafolios para registrar su desarrollo académico o vocacional.
- Los informes pueden incluir desde una puntuación única a un perfil , p.e., desde calificaciones a una descripción pormenorizada de aspectos diversos (Birenbaum, 1996).

## 5. Criterios para las nuevas formas de evaluación

Se han formulado un conjunto de criterios para la práctica de nuevas evaluaciones (Birenbaum, 1996; Collins, 1990; Feltovich y col, 1993; Glaser, 1990; Shavelson, 1994).

En primer lugar, se defiende que una buena pregunta tiene que admitir respuestas que requieren la integración de múltiples componentes o teorías. La coherencia del conocimiento, su estructura e interrelaciones son un objetivo de evaluación.

En segundo lugar, unos buenos instrumentos de evaluación requieren múltiples perspectivas y sensibilidad hacia el contexto. Los estudiantes no necesitan saber únicamente el "qué" sino también "cuándo", "dónde" y "cómo". Esto implica que limitarse a las afirmaciones y las respuestas no es suficiente, es necesario preguntar también por mecanismos causales subyacentes.

En tercer lugar, la evaluación de la aplicación del conocimiento a casos reales es la meta principal de las denominadas prácticas de evaluación innovadoras. Esto equivale a dirigir la evaluación, a analizar en qué medida los estudiantes son capaces de aplicar el conocimiento para resolver problemas de la vida real y tomar las decisiones adecuadas. Glaser (1990, p.47) hace referencia a la importancia de la evaluación del "conocimiento utilizable".

En cuarto lugar, Feltovich et al. (1993, p.209) enfatizan el andamiaje para la transferencia como uno de los objetivos principales de una buena práctica de evaluación. Las preguntas o elementos deben examinar en qué medida la comprensión actual proporciona un andamiaje para la comprensión y el análisis de nuevos problemas en el mismo contenido (Segers y col., 1999).

A medida que la cultura del examen se va transformando en una cultura de evaluación (Birenbaum & Dochy, 1996), se debe también intentar cambiar esta cultura en los estudiantes. Será necesario recurrir con mayor frecuencia a una evaluación mucho más formativa para convencer a los estudiantes de que la evaluación tiene dos propósitos principales. El primero es mostrar a los estudiantes sus puntos fuertes, sus debilidades y su proceso de desarrollo, y el segundo es guiar a los estudiantes hacia el logro de sus metas de aprendizaje.

Las investigaciones sugieren que los estudiantes encuentran con frecuencia las nuevas formas de evaluación interesantes y motivadoras. Si bien los estudiantes nunca pierden su interés por las notas, aprenden y actúan de una manera diferente a la que adoptan en cursos donde se emplean los test tradicionales. La investigación en nuevas formas de evaluación ha dado lugar a las siguientes conclusiones (Birenbaum, 1996; Broadfoot, 1986; Dochy, Moerkerke & Martens, 1996; Segers & Dochy, 1996; Topping, 1998; Wilbrink, 1997):

- Los nuevos métodos de evaluación son menos amenazantes para la mayoría de los estudiantes que los exámenes tradicionales y son percibidos como más justos.
- Los estudiantes encuentran que trabajos tales como proyectos, ejercicios en grupo, y portafolios tienen sentido, debido a su autenticidad y su mayor encaje en entornos de aprendizaje potentes.
- Aunque tales evaluaciones apelan más a la motivación interna del estudiante, las calificaciones siguen siendo importantes en la mente de los estudiantes; y
- El cambio a métodos de evaluación diferentes alienta cambios en los métodos y resultados de aprendizaje que pasan de la memorización pura a un aprendizaje real. (Dochy & Moerkerke, 1997)
- La mayoría de los estudios muestran una alta fiabilidad y validez (Topping, 1998)

En lo anterior hemos esbozado los cambios en la sociedad del aprendizaje y los cambios en la cultura de la evaluación. Esta era de la evaluación nos impone pensar en nuevos criterios de evaluación.

Sin embargo, no deberíamos pecar de optimistas en lo que se refiere a la implementación en la práctica de la cultura de la evaluación. Desafortunadamente, profesores y maestros aún continúan considerando la evaluación como una tarea que se realiza "cuando todo ha acabado". Los responsables educativos, como hemos comprobado, continúan pensando que pueden poner en marcha entornos de aprendizaje de orientación constructivista (también denominados orientados hacia el estudiante) sin necesidad de adaptar las evaluaciones. Una situación en que un aprendizaje orientado a los estudiantes va mano a mano con los exámenes tradicionales (con frecuencia dirigidos hacia la reproducción del conocimiento) nos conduce a lo que en otro momento hemos denominado la "*profecía autodisuelta*". Esta hipótesis afirma que la innovación educacional se esfumará cuando la evaluación no sea congruente con la enseñanza. Un estudiante que trabaje activamente en pequeño grupo en un entorno de aprendizaje basado en problemas para aprender un conocimiento científico básico, para ser capaz de aplicar este conocimiento y aprender diversas habilidades académicas y científicas, seguramente se asustará cuando si finalmente recibe un examen que le exija principalmente reproducir el manual. Como "*homo economicus*", concluirá que es mejor quedarse en casa y memorizar el libro. En los últimos años, he visto varias instituciones en las que esta profecía de autodisolución llega a ser una realidad y los estudiantes forzaron una vuelta a las antiguas prácticas. La evaluación es la piedra angular de la innovación educativa.

En primer lugar, se expondrán las características de los entornos de aprendizaje basados en problemas para la enseñanza además de para la evaluación. En segundo lugar, se discutirán los criterios de calidad para los nuevos modelos de evaluación. En tercer lugar, se presentan dos estudios de casos, el de Maastricht y el de Lovaina. Para

ambos casos se resumirá el sistema de evaluación. Las prácticas de evaluación se relacionarán con las percepciones que los estudiantes tienen de un conjunto de indicadores del entorno de aprendizaje.

## 6. Entornos de aprendizaje basados en problemas

Una de las metas del aprendizaje basado en problemas (ABP) es la de formar estudiantes capaces de analizar y resolver problemas (Barrows, 1986). Por tanto, el proceso de aprendizaje se inicia y se guía por una secuencia de tareas problema variadas que cubren el contenido de una materia . Durante los sucesivos años de estudio, estos problemas se van haciendo más complejos e incluyen diversas actividades que deben ser realizadas por los estudiantes, desde redactar una recomendación dirigida a un empresario a discutir la propuesta en un contexto real con un directivo de una determinada empresa. Trabajando en pequeño grupo (10 -12 estudiantes), guiados por un tutor, los estudiantes analizan el problema presentado y discuten los aspectos relevantes del mismo. Formulan un conjunto de objetivos de aprendizaje basándose en sus hipótesis sobre posibles formas de entenderlo y resolverlo. Estos objetivos son el punto de partida de los estudiantes para aprender los temas en los libros de estudio. En la siguiente sesión de grupo, se presentan y se evalúan los hallazgos de las actividades de autoestudio. El diseño del proceso de enseñanza en el ABP se parece mucho a las características de los entornos de aprendizaje potentes descritos por DeCorte (1990).

Para hacer las prácticas de evaluación congruentes con los principios y la práctica de la instrucción y teniendo en cuenta los principios de la evaluación alternativa mencionada anteriormente, podemos formular los siguientes criterios para una evaluación adecuada:

1. Lo esencial para el ABP es que los estudiantes aprendan analizando y resolviendo problemas representativos de los problemas en los que van a tener que aplicar su conocimiento en el futuro. Consecuentemente, un sistema de evaluación válido evalúa las competencias con un instrumento basado en la vida real, es decir, mediante problemas auténticos (Shavelson, Gao & Baxter, 1996).
2. El estudiante se enfrenta con problemas nuevos, se le pide que transfiera sus conocimiento y sus habilidades y que demuestre que comprende la influencia de factores contextuales en el análisis de un problema, así como en su resolución ;
3. Las tareas de evaluación de análisis de los problemas piden a los estudiantes que argumenten sus ideas sobre la base de perspectivas relevantes diversas;
4. La evaluación de la aplicación del conocimiento cuando se resuelven problemas es el aspecto principal de este enfoque. Por consiguiente, los componentes de la prueba requieren que el estudiante aplique su conocimiento a situaciones de resolución de problemas importantes comunes (Swanson, Case & Van der Vleuten, 1991). Puesto que un nivel suficiente de conocimiento en un dominio específico es un determinante de la resolución de problemas efectiva, la evaluación de la coherencia del conocimiento de los estudiantes sirve al menos como información básica. Dochy (1992) define los perfiles de conocimiento como "un gráfico de puntuaciones en bruto o estandarizadas que describen a un grupo o un individuo en base a ciertos parámetros" (p.143). Éstos indican los puntos fuertes y las debilidades existentes en la base de los conocimientos del estudiante. La investigación ha mostrado que estos perfiles de conocimientos pueden considerarse determinantes básicos de los resultados académicos y pueden identificar con exactitud déficit específicos que contribuyen de manera significativa a unos resultados bajos (Letteri, 1980). En la situación actual, esto implica que los ítems que evalúan el conocimiento tienen que indicar las debilidades existentes en la base del conocimiento del estudiante; ¿es capaz de definir o describir los conceptos centrales de la materia y entender sus interrelaciones?. Esta información mejora el aprendizaje futuro de los estudiantes para obtener la base de conocimientos necesarios para abordar problemas. Por tanto, la adquisición y aplicación del conocimiento son elementos complementarios en un currículum basado en problemas;
5. Los ítems de la prueba exigen más que un conocimiento de conceptos aislados. Se enfatiza la evaluación del conocimiento integrado, que requiere la organización de ideas y conceptos relevantes. Puesto que los problemas de la vida real son en general multidimensionales, y como tales afectan a disciplinas diferentes dentro de un determinado ámbito de estudio, la evaluación se centra en problemas con estas características

integradoras.

6. La interacción social es una eficaz herramienta para un aprendizaje cognitivo y metacognitivo de calidad. Los estudiantes aprenden a reflexionar de manera crítica sobre su propio aprendizaje y sobre los procesos de aprendizaje y los productos de sus compañeros. Adicionalmente, la habilidad para trabajar de manera efectiva en un equipo es una meta crucial dentro de los entornos de ABP. Por consiguiente, la evaluación de los procesos grupales y de la contribución individual a los mismos son parte del sistema de evaluación.

## 7. Nuevas formas de evaluación: cinco métodos bajo el paraguas de la nueva evaluación

En lo que sigue, nos centraremos en los últimos cinco métodos nuevos de evaluación. En primer lugar proporcionaremos una definición o una breve descripción, seguida de los hallazgos principales obtenidos en la investigación.

### 7.1. Evaluación por portafolios

La evaluación por portafolios es un método que favorece la consecución de los objetivos de la nueva evaluación (Birenbaum, 1996). En el pasado, varias profesiones han empleado portafolios para demostrar sus competencias, pero sólo recientemente el portafolios se ha introducido como método de evaluación educativo. (Arter & Spandel, 1992; Paulson, Paulson & Meyer, 1991; Shulman, 1988).

#### *¿Qué es la evaluación por portafolios?*

Arter y Spandel (1992, p.36) definen la evaluación por portafolio como "una colección cuidadosa que reúne trabajos del estudiante y cuenta la historia de sus esfuerzos, su progreso, o sus logros en determinadas áreas. Esta colección está basada en las decisiones del estudiante sobre la selección del contenido del portafolios; las pautas para la selección; los criterios para juzgar el mérito; y la evidencia de autorreflexión".

El portafolio es una "herramienta para el aprendizaje" que integra la enseñanza y la evaluación y se emplea para una evaluación longitudinal de procesos y productos. Propone el diálogo y la cooperación entre los estudiantes y el profesor, en los que los dos colaboran en la evaluación de los resultados y el futuro desarrollo del estudiante con relación a una materia. Además, el portafolio constituye un método que refleja de forma comprensiva, que cuenta la historia del estudiante como aprendiz, señalando sus puntos fuertes y sus debilidades. La elaboración de portafolios anima a los estudiantes a participar y a tomar la responsabilidad de su aprendizaje (Birenbaum, 1996). También proporciona una idea sobre el conocimiento previo del estudiante, a través de la cual la evaluación puede vincularse con ese conocimiento (Dochy, Segers & Buehl, 1999). Finalmente las habilidades y las estrategias de aprendizaje que los estudiantes desarrollarán para construir sus portafolios (búsqueda de datos, organización de los análisis de datos, interpretación y reflexión) son competencias fundamentales en la idea del aprendizaje a lo largo de la vida (Courts & McInerney, 1993; Dochy & Dierick, en prensa).

## Resultados de la investigación

Las investigaciones que se han realizado sobre la influencia de la evaluación por portafolio en las clases, señalan que éstos mejoran la autoevaluación por parte del estudiante (Paulson, Paulson & Meyer, 1991; Tierney, Carter & Desai, 1991) e invitan a la reflexión (Calfee & Perfumo, 1996). Darling-Hammond, Añez y Falk (1995) señalan el impacto positivo del portafolios tanto en la enseñanza como en el aprendizaje de los estudiantes. Además, se ha encontrado que los portafolios son buenos predictores de una ejecución de calidad en el puesto de trabajo.

En contraste con esto, Gentile (1992) correlacionó las puntuaciones de los portafolios con evaluaciones similares en el campo de las matemáticas y la escritura ,y encontró correlaciones débiles. Por otra parte, las relaciones entre los portafolios de los estudiantes realizados en la escuela y las puntuaciones obtenidas en las Evaluaciones Nacionales del Progreso Educativo no eran superiores a las esperadas por azar: las correlaciones variaron desde .13 a .38. Herman, Gearhart y Baker (1993) no encontraron tampoco una correlación significativa entre puntuaciones derivadas de portafolios y una evaluación escrita estándar. Los elementos de los portafolios correlacionaron entre .13 a 0.31 con la evaluación estándar.

En el trabajo de la Corporación RAND de Vermont, participaron evaluadores especialmente formados, no los profesores de los estudiantes que habían elaborado los portafolios, que juzgaron de forma independiente una muestra de estudiantes de cuarto y octavo curso. La correlación varió entre evaluadores (basándose en los datos del segundo año de la implementación de esta forma de evaluación en todo el estado) entre .40 a .63, dependiendo del número de componentes del portafolio y del número de dimensiones empleadas por los evaluadores (Koretz, McCaffrey, Kelin, Bell & Stecher, 1993). De acuerdo con Gredler (1995) la ausencia de información válida y fiable es en el mejor de los casos, problemática.

Otros investigadores informan sobre una fiabilidad más elevada entre evaluadores. En un análisis de la fiabilidad de los portafolios en la escuela y el instituto realizado en el distrito de la escuela pública del distrito de Pittsburgh emplearon evaluaciones que no eran las utilizadas por los profesores y los alumnos. La fiabilidad varió de .58 a .87 dependiendo de la dimensión examinada y el nivel académico (LeMahieu, Gitomer & Eres, 1995). En un estudio más modesto, se emplearon datos de portafolios recogidos en tres clases de escuela elemental. La fiabilidad varió desde .76 a .94, dependiendo de si los juicios de los resultados de los estudiantes se basaban en deberes escritos estándar, muestras individuales del trabajo del estudiante, o el portafolios como un todo (Herman, Gearhart & Baker, 1993). En un estudio sobre el acuerdo interjueces en la evaluación de portafolios de las artes relacionadas con el lenguaje en los primeros cursos de la escuela de Rochester (Nueva York), los profesores de la clase y un observador externo evaluaron el portafolios. La fiabilidad en lectura osciló de .58 para el primer curso a .77 para preescolar. La fiabilidad para las notas de escritura osciló entre .68 para preescolar y .73 para el primer curso (Supovitz, Macqowan & Slattery, 1997).

## 7.2 *La evaluación reconceptualizada en la práctica: El uso de casos para una Prueba General*

La Prueba General (*OverAll Test u OAT*) es un ejemplo de la evaluación reconceptualizada en la práctica educativa que está basada en las ideas actuales (Dochy & Moerkerke, 1997). Esta nueva forma de evaluación se emplea en la actualidad principalmente en Facultades de Económicas y Ciencias Empresariales donde se creó.

*¿Qué es una Prueba General?*

La Prueba General evalúa en qué medida los estudiantes son capaces de analizar problemas y contribuir a su solución aplicando las herramientas relevantes. Adicionalmente, mide si los estudiantes son capaces de identificar y seleccionar los instrumentos más relevantes para resolver los problemas presentados, es decir, si saben “cuándo y dónde” (Segers, 1996<sup>a</sup>; Segers, 1997).

La Prueba General se utiliza dentro del diseño curricular de la siguiente forma. Tras dos períodos o bloques instruccionales los estudiantes tienen dos semanas libres para el estudio independiente. Durante esas dos semanas, trabajan de un manual de estudio que reciben al comienzo de dicho periodo. Este manual contiene información sobre los objetivos principales de la Prueba General, las partes del programa son relevantes para el estudio del material presentado en el manual, un ejemplo de un caso con las preguntas relevantes, la información práctica (organizativa) relevante y, finalmente, un conjunto de casos. Las características de dichos casos varían. Puede ser una descripción de las innovaciones o los problemas de una empresa nacional o internacional que han aparecido en los periódicos, una revista o el estudio de un caso. En otros casos, se trata de artículos de revistas científicas, que contienen consideraciones teóricas de un científico, el informe de un estudio o comentarios de una teoría o modelo. Durante el periodo de estudio se espera que los estudiantes aplicarán el conocimiento que han adquirido durante las semanas anteriores para explicar los problemas nuevos, complejos que se les presenta en el conjunto de los artículos. Se les pide, como al leer los casos, que traten de explicar (sin el estímulo de un tutor) las ideas y teorías descritas en dichos artículos relacionándolas con los conocimientos que previamente han adquirido. Esto se denomina con frecuencia “auto-explicación” (Chi, de Leeuw, Chiu & LaVancher, 1992). En resumen, el periodo de estudio plantea la posibilidad de que los estudiantes analicen y sintetizan los problemas económicos tal y como han sido abordados en los grupos de trabajo. Por consiguiente, el manual de estudio les ofrece un conjunto de nuevos problemas descritos en un conjunto de casos (Segers y col. 1999).

La Prueba General es una evaluación de papel y lápiz. Las preguntas que contiene se basan en los diferentes artículos que han estudiado en casa y en los casos que se distribuyen con ocasión de la prueba (Segers, 1996b). Combina dos formatos de ítems: preguntas de verdadero / falso que contiene la opción “no sé” y preguntas abiertas. Los ítems de verdadero / falso se emplean para medir si un estudiante puede aplicar en una nueva situación el conocimiento adquirido, si puede usar un concepto abstracto relevante para la “vida real de un economista” en una situación compleja y específica. Por el contrario, las preguntas abiertas requieren elaboración, que no puede evaluarse fácilmente mediante un formato de respuesta múltiple. A los estudiantes se les pide que analicen un problema nuevo, por ejemplo, que extraigan dos posibles perfiles de escenarios una matriz de correlaciones y que evalúen los resultados de dichos perfiles. Las preguntas abiertas y los subtests de verdadero / falso tienen el mismo peso.

La Prueba General combina la intención de los profesores de simular situaciones del mundo real con la utilización de problemas auténticos, así como un carácter de libro abierto. Esto significa que los estudiantes pueden traer todo el material de estudio que crean que van a necesitar. Como en el mundo real, tienen todas las fuentes disponibles (Segers y col. 1999). Los estudiantes tienen que ser capaces de seleccionar los materiales y los instrumentos apropiados para resolver las cuestiones. Sin embargo, si son capaces de emplearlos de una manera interpretativa, no serán capaces de analizar y resolver los problemas que se les proponen (Feller, 1994).

## Hallazgos en investigación

Las preguntas de la Prueba General hacen referencia a entre siete o doce casos o artículos que describen uno o más problemas relacionados con la economía. La elección de este número de casos se basa en el hecho de que si la amplitud de la muestra es limitada, la generalización de las puntuaciones puede ser deficiente debido a la especificidad del contenido. Estos hallazgos se confirman por los resultados en un estudio piloto con la Prueba General (Segers, Tempelaar, Keizer, Schijns, Vaerssen & Van Mourik, 1991; 1992). La mayor parte de la varianza se debió al efecto de la interacción de personas y casos (35% para la parte abierta y el 65% para el test de verdadero / falso). Esto significa que los estudiantes con mejores resultados en un caso no obtienen necesariamente los mejores resultados en los otros casos. Esto significa además, que un caso tiene un bajo valor predictivo para los

otros casos. Los resultados sugieren que para una Prueba General con 12 casos, el coeficiente de generalidad es 0.67 (Segers y col. 1999). Segers investigó la equidad de la Prueba General así como la medida en que esta prueba mide la capacidad de los estudiantes para recurrir a una red conceptual para analizar problemas auténticos. El estudio sugiere un importante grado de solapamiento entre el currículum formal y el operacional en términos de los conceptos estudiados así como de en términos de los niveles de dominio que se pretendían y se alcanzaron. Igualmente, se encontró una congruencia suficiente entre la práctica de evaluación en términos de metas evaluadas y el currículum formal y operacional. Esto significa que, aún cuando los estudiantes disfrutan de una considerable libertad dentro del programa, la Prueba General es justa para los estudiantes. Adicionalmente, los resultados indican que es posible evaluar la capacidad de los estudiantes para resolver problemas con la Prueba General. Sin duda, la evidencia muestra que la Prueba General presenta un buen nivel de congruencia entre la enseñanza en un entorno de aprendizaje estimulante y la evaluación. Por tanto, es un instrumento digno de consideración, ya que su aplicación no es muy diferente de las formas tradicionales de evaluación utilizadas por los profesores (pruebas de elección múltiple y preguntas abiertas).

### 7.3 *Evaluaciones en las que los estudiantes desempeñan el papel de evaluadores*

En esta sección, discutiremos los principales resultados de la investigación sobre nuevas formas de evaluación como la autoevaluación, evaluación de iguales y la co-evaluación (Dochy, Segers y Sluijsmans, 1999). El uso de estas nuevas formas de evaluación es consistente con las demandas sociales de formar aprendices a lo largo de toda la vida, capaces de reflexionar de forma continua sobre su conducta y sobre sus propios procesos de aprendizaje (Moerkerke, 1996).

#### 7.3.1 *Auto-evaluación*

##### *¿Qué es la autoevaluación?*

La autoevaluación hace referencia a la participación de los aprendices en la realización de juicios acerca de su propio aprendizaje, particularmente acerca de sus logros y de los resultados de su aprendizaje (Boud & Falchikov, 1989). La autoevaluación no es una técnica nueva, sino un medio de potenciar el papel de los estudiantes como participantes activos en su propio aprendizaje (Boud, 1995). Se utiliza con frecuencia para evaluaciones formativas con el objetivo de fomentar la reflexión sobre los propios procesos de aprendizaje y sus resultados.

#### **Resultados de la investigación**

Boud y Falchikov (1989) analizaron los estudios publicados entre 1932 y 1988, sobre la comparación de las autoevaluaciones de los estudiantes con las evaluaciones de los profesores. Tanto las sobrevaloraciones como las infravaloraciones de los estudiantes se relacionaron con sus habilidades. Los resultados indicaron que los buenos estudiantes tienden a subestimarse y que los estudiantes más flojos se sobrevaloraban. Igualmente los estudiantes de las clases superiores podían predecir mejor su rendimiento que los estudiantes de niveles más bajos.

Algunos estudios confirman que la habilidad de los estudiantes para evaluarse a sí mismos mejora con la retroalimentación a través del tiempo (Birenbaum y col.1996; Bound y col. 1989; Griffee, 1995). Además, las interpretaciones de los estudiantes no sólo dependen de la forma del proceso de evaluación, sino también en la manera en cómo esas tareas se sitúan dentro del contexto total de la materia y dentro de su experiencia global de la vida académica.

Longhurst y Norton (1997) diseñaron un estudio para investigar si los estudiantes de segundo año de Psicología serían capaces de evaluar sus propias producciones analizando también si los estudiantes entendían qué significa adoptar un enfoque profundo en la elaboración de sus trabajos. La nota de los estudiantes se comparó con la nota de los tutores. Los tutores también calificaron los trabajos en base a un conjunto de criterios de procesamiento profundo. Los resultados mostraron que las notas de los tutores correlacionaron de manera significativa con el conjunto de criterios de procesamiento profundo ( $r$  entre .69 y .88). Además, se comprobó que existía una correlación positiva entre las notas de los estudiantes y las notas de los profesores ( $r$  .43). Los resultados indican que, en conjunto, los estudiantes fueron precisos en la evaluación de sus propios trabajos pero menos precisos en la evaluación del procesamiento profundo. Los estudiantes menos motivados y los menos capaces parecían tener menos clara la evaluación de los criterios individuales.

Zoller y Ben-Chaim (1997) investigaron la habilidad para autoevaluarse de estudiantes de biología matriculados en un programa universitario de cuatro años, con respecto a las Habilidades Cognitivas de Orden superior así como su confianza en la autoevaluación. Los resultados señalaron que los estudiantes consideraban que tenían suficientes conocimientos. Asimismo mostraron que el 75% de los estudiantes pensaban que eran capaces de autoevaluarse y de evaluar a sus compañeros. Zoller y Ben-Chaim encontraron una discrepancia entre las evaluaciones de los estudiantes y las evaluaciones de los profesores que explicaron en términos de la falta de integración entre la evaluación y el aprendizaje en la actual enseñanza de la ciencia.

Los resultados de investigación realizados por Hassmén indican que los estudiantes que realizan una autoevaluación mientras aprenden obtienen un porcentaje superior de respuestas correctas en pruebas de aprendizaje que aquellos que aprenden sin autoevaluarse.

Dochy et al.(1999) concluyeron a partir de su revisión bibliográfica, que los estudiantes que practican la autoevaluación tienden a puntuar más alto en los exámenes. La autoevaluación, empleada en la mayoría de los casos para fomentar el aprendizaje de habilidades y capacidades, lleva a una mayor reflexión sobre el propio trabajo, a unos criterios más elevados, a una mayor responsabilidad por su propio aprendizaje y al incremento de la comprensión en la resolución de problemas. La exactitud de la autoevaluación mejora con el tiempo especialmente cuando los profesores proporcionan a los estudiantes información sobre sus autoevaluaciones.

### 7.3.2 *Evaluación de iguales*

#### *¿Qué es la evaluación de pares o iguales?*

Falchikov (1995) define evaluación de iguales como el proceso por el que grupos de individuos evalúan a sus compañeros. Este ejercicio puede o no implicar una discusión o acuerdo previo sobre los criterios. Los criterios pueden referirse a dominio del contenido, dominio de habilidades científicas o académicas, dominio de habilidades sociales y de comunicación, etc.

### **Resultados de investigación**

Dancer y Dancer (1992) indican que las investigaciones no han mostrado la validez de la evaluación por iguales. Los pares tienden a realizar evaluaciones basadas en la uniformidad, la raza y la amistad en ausencia de un entrenamiento intensivo en la evaluación de iguales. Por el contrario, Topping (1998) revisó 31 estudios y concluyó que 18 de esos estudios mostraban una alta validez y fiabilidad en diversos campos y sólo 7 estudios encontraron que la validez y la fiabilidad eran inaceptablemente bajas.

Conway, Kember, Sivan y Wu (1993) investigaron la equidad de la evaluación de iguales examinando las vías por las que los estudiantes podían asignar las notas individuales, reflejando el esfuerzo personal, para proyectos grupales. Conway et al., adoptaron los procedimientos sugeridos por Goldfinch y Raeside (1990) y los simplificaron. Los resultados, usando este método para calcular un peso individual, mostraron que los estudiantes percibían la evaluación de iguales como un buen método y suficientemente justo. Los estudiantes pensaban también que debían jugar una parte en la evaluación para que los resultados de la evaluación fueran más objetivos.

Los estudios que investigan la exactitud de la evaluación de iguales muestran resultados diversos. Oldfield y Macalpine (1995) investigaron la competencia de los estudiantes al realizar estas evaluaciones. Sus resultados muestran altas correlaciones entre las notas de los estudiantes y las calificaciones que los profesores realizaban de sus trabajos y presentaciones. Fry (1990) describe un estudio en el que el tutor utilizó las calificaciones de iguales. El acuerdo entre las notas del tutor y las notas de los estudiantes era generalmente muy alto. Los hallazgos de Fry fueron confirmados por Rushton, Ramsey y Rada (1993), que desarrollaron una herramienta computerizada de evaluación por pares.

Los resultados de un estudio de Orsmond, Merry y Reiling (1996) sobre la exactitud de la evaluación de iguales son menos positivos. Después de que los estudiantes calificaran los posters elaborados por sus compañeros el tutor también los puntuó sin conocer las puntuaciones que los estudiantes habían asignado. Orsmond et al (1996) encontraron que había poco acuerdo entre los estudiantes y el tutor, ya que la correlación fue de .54. Estos resultados están en la línea de los resultados encontrados por Stefani (1994).

Se ha informado de distintos efectos positivos de la evaluación de iguales. Orsmond y col. (1996) encontraron que los estudiantes disfrutaban llevando a cabo la evaluación y que ésta era beneficiosa para su aprendizaje. Los estudiantes contestaron un cuestionario en el que el 76% indicaban que "la evaluación de iguales les había hecho pensar más, y trabajar de un modo más estructurado" (p.243). Keaten y Richardson (1992) también afirmaron que dicha evaluación fomentaba una mejor apreciación de las aportaciones individuales dentro del grupo así como de las relaciones interpersonales en la clase.

Williams (1992) encontró que la gran mayoría de estudiantes mostraron beneficios con esta evaluación. Sin embargo los estudiantes encontraron que criticar a sus amigos podía ser difícil. Los estudiantes encontraron además que las evaluaciones de iguales eran difíciles o no deseables cuando se establecían desde un principio las directrices de la evaluación. Los dos hallazgos más importantes del estudio de Williams (1992) fueron: (1) que a los estudiantes les gustaba tener una participación más activa en lo que respecta a su aprendizaje y su evaluación y (2) que los estudiantes necesitan ser guiados y entrenados en este nuevo papel antes de que pueda recurrirse a él. Cheng y Warren (1997) concluyeron también a partir de su estudio que los estudiantes necesitaban ser entrenados de manera sistemática y comprensiva en cómo evaluar a sus iguales y en el establecimiento de criterios.

Dochy et al. (1999) llegan a la conclusión, a partir de su revisión de la literatura, de que la evaluación por pares puede ser valiosa tanto como evaluación formativa como método de evaluación sumativa (introduciendo un factor de corrección para los productos grupales) y, por lo tanto, como parte del proceso de aprendizaje ya que con ella, los estudiantes se sienten más involucrados en el proceso de aprendizaje y de evaluación. La evaluación de iguales se considera como suficientemente justa y exacta.

### 7.3.3. *Co-evaluación*

### *¿Qué es la co-evaluación?*

De acuerdo con Hall (1995) la co-evaluación, o la participación de estudiantes del profesorado en el proceso de evaluación, es una manera de proporcionar a los estudiantes una oportunidad de evaluarse ellos mismos a la vez que permite al profesorado mantener el control necesario sobre la evaluación final.

Somervell (1993) considera la evaluación colaborativa como un proceso de enseñanza y aprendizaje en el que el estudiante y el instructor deben ponerse de acuerdo para clarificar objetivos y criterios. En este caso, el estudiante no es necesariamente responsable de la evaluación, pero colabora en el proceso de determinar lo que debe ser evaluado y tal vez, por quién será evaluado.

Pain, Bull y Brna (1996) argumentan que el término “evaluación colaborativa” puede ser aplicado a un evaluador y a un evaluado que trabajan juntos para llegar a una visión consensuada sobre el conocimiento del estudiante. Es una verdadera colaboración en la medida en que ambas partes trabajan para alcanzar el objetivo compartido de proporcionar una evaluación consensuada del conocimiento del estudiante. Esta colaboración implica que ambas partes negocian los detalles de la evaluación y discutan cualquier malentendido que exista, y es consistente con un enfoque de la evaluación menos basado en la confrontación que se valora cada vez más y que se desarrolla a partir de la relación existente entre evaluador y evaluado.

## **Resultados de las investigaciones**

En el estudio de Hall (1995) los estudiantes y el profesorado decidieron conjuntamente los criterios de evaluación. Se identificaron tres propuestas de co-evaluación. La primera consistió en ayudar al estudiante de educación a realizar paulatinamente su cambio de papel de estudiante a profesor; la segunda consistió en proporcionar una reflexión sobre el proceso de evaluación que podía resultarles útil en la evaluación de sus propios estudiantes, y la tercera consistió en proporcionar el desarrollo de una habilidad que lleve a los estudiantes a la auto-evaluación. El proceso incluía la asignación de una hoja con dos caras para cada tarea. Por detrás de esta hoja los estudiantes podían dar su propia auto-evaluación que después pasaban al profesor. El profesor a continuación, utilizaba esta hoja para anotar su evaluación del trabajo de los estudiantes. Finalmente el profesor le daba la vuelta a la hoja para ver si el estudiante había elegido o no ofrecer su propia evaluación en la otra cara. Los resultados fueron que, generalmente, la calificación del profesor era más alta que la calificación de los estudiantes.

Orpen (1982) estudió a 21 estudiantes de un curso sobre comportamiento organizacional y a 21 estudiantes de un curso de filosofía política a los que se les elaboraron un trabajo. Se les informó que “sus trabajos serían evaluados más adelante por cinco profesores, y que su calificación final sería la media de las notas que recibieran de sus compañeros y de los profesores” (p.568). Las notas se asignaron de acuerdo al siguiente criterio: (1) cobertura del material relevante, (2) coherencia y fuerza del argumento subyacente, (3) fluidez y claridad de expresión. Los resultados indicaron que no había diferencia entre los profesores y los estudiantes en las que asignaron en promedio notas, ni en la varianza de estas notas, ni en el acuerdo entre estas calificaciones ni en la relación entre estas notas y los resultados de los estudiantes en los exámenes de fin de curso.

Los estudios que examinaron la exactitud de estas evaluaciones muestran la importancia de establecer criterios previos, conjuntamente por el grupo y el profesor o por estudiantes de manera independiente (Stefani, 1994; Falchikov, 1986; Kwan & Leung, 1996). Horgan, Bol y Hacker (1997), además, hacen hincapié en los efectos del tiempo y del entrenamiento.

## 8. Conclusiones

En lo que antecede hemos revisado las nuevas demandas que la sociedad formula a las instituciones docentes, así como los nuevos métodos de enseñanza y de evaluación que la satisfacción de estas demandas implica. Estos métodos innovadores enfatizan la adopción de un papel activo por parte del estudiante, tanto en lo que se refiere a su responsabilización para alcanzar los logros académicos que se proponen, así como en el desarrollo y utilización de criterios de evaluación que de hecho constituyen, por sí mismos, objetivos genuinos del proceso educativo.

## Referencias

- Albanese, M.A., & Mitchell, S. (1993). Problem-based learning: A review of literature on its outcomes and implementation issues. *Academic Medicine*, p. 68, pp. 52-81.
- Arter, J.A., & Spandel, V. (1992). Using portfolios of student work in instruction and assessment. *Educational Measurement: Issues and Practice*, 11 (1), pp. 36-44.
- Barrows, H.S. (1986). Problem-based learning in medicine and beyond: A brief overview. En L. Wilkerson & W.H. Gijselaers (Eds.): *New Directions for Teaching and Learning*, Nº 68, pp. 3-11. San Francisco: Jossey-Bass Publishers.
- Beckwith, J.B. (1991). Approaches to learning, their context and relationship to assessment performance. *Higher Education*, 22, pp. 17-30.
- Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. En M. Birenbaum & F.J.R.C. Dochy (Eds.): *Alternatives in Assessment of Achievements, Learning Processes and Prior Knowledge*, pp. 3-30. Boston: Kluwer Academic Publishers.
- Boud, D. (1995) Enhancing Learning through Self-assessment. En Kogan Page (Ed.). London: Philadelphia.
- Boud, D., & Falchikov, N. (1989). Quantitative Studies of Self-assessment in Higher Education: a Critical Analysis of Findings. *Higher Education*, 18(5), pp. 529-549.
- Broadfoot, P.M. (1986). Profiles and records of achievement: A review of issues and practice. London: Holt, Rinehart & Winston.
- Calfee, R., & Perfumo, P. (1996). Writing portfolios in the classroom; Policies and practice, promise and peril. Mahwah: NJ:Lawrence Erlbaum Associates.
- Campione, J.C., & Brown, A.L. (1990). Guided learning and transfer : Implications for approaches to assessment. In N. Frederiksen, R. Glaser, AA. Lesgold & M.G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 141-172). Hillsdale: NJ: Lawrence Erlbaum Associates.
- Cheng, W., & Warren, M. (1997). Having Second Thoughts: student perceptions before and after a peer assessment exercise. *Studies in Higher Education*, 22, 233-239.
- Chi, M.T.H., Leeuw de, N., Chiu M. & LaVancher, C. (1992). Self-explanations Improve Learning. Paper presented at the NATO Advanced Study Institute on 'Psychological and Educational Foundations of Technology-Based Learning Environments' at Kolymbari. Crete, July 1992.
- Collins, A. (1990). Reformulating Testing to Measure Learning and Thinking. In N. Frederiksen, R. Glaser, A. Lesgold, & M.G. Shafto, *Diagnostic Monitoring of Skill and Knowledge Acquisition*, pp. 75-87. Hillsdale: N.J.: Lawrence Erlbaum Associates.
- Conway, R., Kember, D., Sivan, A., & Wu, M. (1993). Peer Assessment of an Individual's Contribution to a Group Project. *Assessment and Evaluation in Higher Education*, 18(1), pp. 45-56.
- Courts, P.L., & McIreney, K.H. (1993). *Assessment in higher education. Politics, pedagogy, and portfolios*. London: Westport, Connecticut.
- Dancer, W.T., & Dancer, J. (1992). Peer Rating in Higher Education. *Journal of Education for Business*, 67(5), pp. 306-309.
- Darling-Hammond, L. Ancess, J. & Falk, B. (1995). *Authentic assessment in action*. New York: Teachers College Press.
- De Corte, E. (1990). A State-of-the-art of research on learning and teaching. Keynotelecture presented at the first European Conference on the First Year Experience in Higher Education. Aalborg University, Denmark, pp. 23-25.
- De Rijke, R., & Dochy, F. (1995). Assessment centres in onderwijs, opleiding en HRM. Utrecht: Lemma.
- Dochy, F. (1994). Prior knowledge and learning. En T. Husén & T.N. Postlethwaite (Eds.), *International encyclopedia of education*, Second edition, pp. 4698-4702. Oxford/New York:Pergamon Press.
- Dochy, F., & Dierick, S. (in press). Een structureel beleid voor toetsing en assessment in constructiegericht onderwijs (CGO): Kern voor het slagen van universitaire onderwijsvernieuwing. *Tijdschrift voor Onderwijsrecht en -beleid*.
- Dochy, F., & Moerkerke, G. (1997). The present, the past and the future of achievement testing and

- performance assessment. *International Journal of Educational Research*, p. 27, p. 5, pp. 415-432.
- Dochy, F., Moerkerke, G., & Martens, R. (1996). Integrating assessment, learning and instruction: assessment of domain-specific and domain-transcending prior knowledge and progress. *Studies in Educational Evaluation*, 22 (4), pp. 123-136.
- Dochy, F., Segers, M., & Buehl, M. (1999). The Relation Between Assessment Practices and Outcomes of Studies: The Case of Research on Prior Knowledge. *Review of Educational Research*, Summer 1999, Vol. 69, No. 2, pp. 147-188.
- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: a review. *Studies in Higher Education*, 24 (3), pp. 331-350.
- Dochy, F.J.R.C., & McDowell, L. (1997). Assessment as a tool for learning. *Studies in Educational Evaluation*, 23 (4), pp. 279-298.
- Dochy, F.J.R.C., & Moerkerke, G. (1997). Assessment as a major influence on learning and instruction. *International Journal of Educational Research*, 27 (5). Pp. 415-432.
- English, F.W. (1992). Deciding what to teach and test. Newbury Park California: Sage.
- Falchikov, N. & Boud, D. (1989). Student Self-assessment in Higher Education: A Meta-Analysis. *Review of Educational Research*, 59(4), pp. 395-430.
- Feller, M. (1994). Open-book testing and education for the future. *Studies in Educational Evaluation*, 20, pp. 235-238.
- Feltovich, P.J., Spiro, R.J. & Coulson, R.L. (1993). Learning, Teaching, and Testing for Complex Conceptual Understanding. En N. Frederiksen, R.J. Mislevy & I.I. Bejar (eds.), *Test theory for a New Generation of Tests*. Hillsdale, N.J.: Lawrence Erlbaum Associates, Publishers.
- Fry, S.A. (1990). Implementation and Evaluation of Peer Marking in Higher Education. *Assessment and Evaluation in Higher Education*, 15(3), pp. 177-189.
- Gentile, C. (1992). Exploring new methods for collecting students' school-based wiring: NAEP's 1990 portfolio study. Washington, DC: National Center for Educational Statistics.
- Glaser, R. (1990). Toward new models for assessment. *International Journal of Educational Research*, 14, pp. 375-483.
- Glaser, R., & Silver, E. (1994). Assessment, Testing and Instruction: Retrospect and Prospect. *of Research in Education*, 20, pp. 393- 419.
- Glaser, R., Lesgold, A., & Lajoie, S. (1987). Toward a cognitive theory for the measurement of achievement. In R.R. Ronning, J. Glover, J.C. Conoley & J.C. Witt (Eds.), *The influence of cognitive psychology on testing and measurement*, pp. 41-85. Hillsdale, NJ: Lawrence Erlbaum.
- Green, D., Brannigan, C., Mazelan, P., & Giles, L. (1994). Measuring student satisfaction: A method for improving the quality of the student experience. En S. Haselgrave (Ed.), *The student experience*, pp. 100-107. Buckingham, UK: SRHE & Open University Press.
- Griffee, D.T. (1995). *A Longitudinal Study of Student Feedback: Self-assessment, Course Evaluation and Teacher Evaluation*. U.S.: Birmingham, Alabama.
- Hall, K. (1995). Co-assessment: participation of students with staff in the assessment process. A report of Work in Progress. Paper given at the 2nd European Electronic Conference On Assessment And Evaluation, EARLI-AE list European Academic & Research Network (EARN) (<http://listserv.surfnet.nl/archives/earli-ae.html>).
- Hambleton, R.K., & Murphy, E. (1992). A psychometric perspective on authentic measurement. *Applied Measurement in Education*, 5, pp. 1-16.
- Harvey, L. & Green, D. (1994) Quality in Higher Education Project: Employer Satisfaction Summary Report. Birmingham, UK: University of Central England.
- Hassmen, P., Sams, M.R., & Hunt, D.P. (1996). Self-assessment responding and testing methods: effects on performers and observers. *Perceptual and Motor Skills*, 83, pp. 1091-1104.
- Herman, J., Gearhart, M., & Baker, E. (1993). Assessing writing portfolios: Issues in the validity and meaning of scores. *Educational Assessment*, 1 (3), pp. 201-224.
- Horgan, D.D., Bol, L. & Hacker, D. (1997). An examination of the Relationships Among Self, Peer, and Instructor Assessments. Paper presented at the European Association for Research on Learning and

- Instruction, Athens, Greece, 1997.
- Keaten, J.A. & Richardson, M.E. (1992, February). A field investigation of peer assessment as part of the student group grading process. Paper presented at the Western Speech Communication Association Convention, Albuquerque, NM.
- Keeves, J.P. (1994). Methods of assessment in schools. En T. Husén & T.N. Postlethwaite (Eds.), International Encyclopedia of Education, pp. 362-370. Oxford / New York: Pergamon Press.
- Kleinasser, A., Horsch, E., & Tastad, S. (1993, April). Walking the talk: Moving from a testing culture to an assessment culture. Paper presented at the Annual Meeting of the American Educational Research Association. Atlanta, GA.
- Koretz, D., McCaffrey, D., Kelin, S. Bell, R., Stecher, D. (1993). Interim report: The reliability of Vermont portfolio scores in the 1992-93 school year. Santa Monica, CA: RAND Institute on Education and Training, National Center for Research on Evaluation, Standards and Student Testing.
- Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement; Issues and Practice*, 13 (3), pp. 5-16.
- Kwan, K., & Leung, R. (1996). Tutor versus Peer Group Assessment of Student Performance in a Simulation Training Exercise. *Assessment and Evaluation in Higher Education*, 21(3), pp. 205-214.
- LeMahieu, P., Gitomer, D.H., & Eresh, J.T. (1995). Portfolios in large-scale assessment: Difficult but not impossible. *Educational Measurement: Issues and Practices*, 14 (3), pp. 11-16, pp. 25-28.
- Letteri, C.A. (1980). Cognitive profile: basic determinant of academic achievement. *The Journal of Educational Research*, 4, 195-198. Letteri, C.A., Kuntz, S.W. (1982). Cognitive profiles: examining self-planned learning and thinking styles. Paper presented at the Annual American Educational Research Association Meeting, New York City, March, pp. 19-23.
- Lohman, D.F. (1993). Teaching and testing to develop fluid abilities. *Educational Researcher*, 22, pp. 12-23.
- Longhurst, N., & Norton, L.S. (1997). Self-assessment in Coursework Essays. *Studies in Educational Evaluation*, 23(4), pp. 319-330.
- Magone, M.E., Cai, J., Silver, E.A. & Wang, N. (1994). Validating the cognitive complexity and content quality of a mathematics performance assessment. *International Journal of Educational Research*, 21, (4), pp. 317-340.
- Masters, G., & Mislevy, R.J. (1993). New views of student learning : Implications for educational measurement. En N. Frederiksen, R.J. Mislevy, & I. Bejar (Eds.), *Test theory for a new generation of tests*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Mislevy, R.J. (1996) Test theory reconceived. *Journal of Educational Measurement*, 33 (4), pp. 379-416.
- Moerkerke, G. (1996). Assessment for flexible learning. Utrecht: Lemma.
- Nitko, A. (1995) Curriculum-based continuous assessment: a framework for concepts, procedures and policy, *Assessment in Education*, 2, pp. 321-337.
- Oldfield, K.A., & Macalpine, J.M.K. (1995). Peer and Self-assessment at the Tertiary Level-An Experiential Report. *Assessment and Evaluation in Higher Education*, 20(1), pp. 125-132.
- Orsmond, P., Merry, S., & Reiling, K. (1996). The importance of Marking Criteria in the Use of Peer Assessment. *Assessment and Evaluation in Higher Education*, 21(3), pp. 239-249.
- Pain, H., Bull, S. & Brna, P. (1996). A student model 'for its own sake' [On-line]. Available: <http://cbl.leeds.ac.uk/~paul/papers/euroaiedpapers96/smpaper/smpaper.html>
- Paulson, F.L., Paulson, P.R., & Meyer, C.A. (1991). What makes a portfolio a portfolio? *Educational Leadership*, 48 (5), pp. 60-63.
- Resnick, L.B. (1987). Learning in school and out. *Educational Researcher*, 16 (9), pp. 13-20.
- Rushton, C., Ramsey, P. & Rada, R. (1993). Peer Assessment in a Collaborative Hypermedia Environment: A Case Study. *Journal of Computer-Based Instruction*, 20(3), pp. 75-80.
- Sambell, K. & McDowell, L. (1997). The Value of Self- and Peer Assessment to the Developing Lifelong Learner. In C. Rust (Ed.), *Improving student learning - improving students as learners*, pp. 56-66. Oxford, UK: Oxford Centre for Staff and Learning Development.
- Schon, D.A. (1987). Educating the Reflective Practitioner: Towards a New Design for Teaching and

- Learning in the Professions. San Francisco: Jossey-Bass.
- Segers, M. S. R. (1997). An alternative for assessing problem-solving skills: The OverAll Test. *Studies in Educational Evaluation*, 23 (4), pp. 373-398.
- Segers, M., Dochy, F., & De Corte, E. (1999). Assessment practices and students' knowledge profiles in a problem-based curriculum. *Learning Environments Research*, 2, pp. 191-213.
- Segers, M.S.R. (1996a). Assessment in a problem-based economic curriculum. In M. Birenbaum & F.J.R.C. Dochy (Eds.), *Alternatives in Assessment of Achievements, Learning Processes and Prior Knowledge*, pp. 201-224. Boston: Kluwer Academic Publishers.
- Segers, M.S.R. (1996b). An Alternative for Assessing Problem Solving Skills: the OverAll Test. Paper presented at the Northumbria Assessment Conference, Newcastle, 4-6 September 1996.
- Segers, M.S.R. (1999). Assessment in student-centered education: does it make a difference? *UNISCENE Newsletter*, 2 (in press).
- Segers, M.S.R., & Dochy, F.J.R.C. (1996). The use of performance indicators for quality assurance in higher education. *Studies in Educational Evaluation*, 22 (2), pp. 115-139.
- Segers, M.S.R., Tempelaar, D., Keizer, P., Schijns, J., Vaessen, E., & Van Mourik, A. (1991). De overall-toets: een eerste experiment met een nieuwe toets vorm. Maastricht: University of Limburg.
- Segers, M.S.R., Tempelaar, D., Keizer, P., Schijns, J., Vaessen, E., & Van Mourik, A. (1992). De overall-toets: een tweede experiment met een nieuwe toetsvorm. Maastricht: university of Limburg.
- Shavelson, R.J. (1994). Guest Editor Preface. *International Journal of Educational Research*, p. 21, pp. 235-237.
- Shavelson, R.J., Xiaohong, G., & Baxter, G. (1996). On the content validity of performance assessments: Centrality of domain-specifications. In M. Birenbaum, & F. Dochy. *Alternatives in assessment of achievements, learning processes and prior knowledge*, pp. 131-142. Boston: Kluwer Academic.
- Shulman, L.S. (1988). A union of insufficiencies: Strategies for teacher assessment in a period of educational reform. *Educational Leadership*, 45, 36-14.
- Snow, R.E., & Lohman, F.D. (1989). Implications of cognitive psychology for educational measurement. En R.L. Linn (Ed.), *Educational measurement* (3rd ed.) pp. 263-331). New York: American Council on Education/Macmillan.
- Somervell, H. (1993). Issues in Assessment, Enterprise and Higher Education: The Case for Self-, Peer and Collaborative Assessment. *Assessment and Evaluation in Higher Education*, 18(3), pp. 221-33.
- Stefani, L.A.J. (1994). Peer, self- and tutor assessment: relative reliabilities. *Studies in Higher Education*, 19(1), pp. 69-75.
- Supovitz, J.A., MacGowan III, A., & Slattery J. (1997). Assessing agreement: An examination of the interrater reliability of portfolio assessment in Rochester, New York. *Educational Assessment*, 4 (3), pp. 237-259.
- Swanson, D.B., Case, S.M., & Van der Vleuten, C.P.M. (1991). Strategies for student Assessment. In D. Boud & G. Feletti (Eds.), *The Challenge of Problem-based Learning* (2º Ed.), pp.269-282. London: Kogan Page.
- Tierney, R.J., Carter, M.A., & Desai, L.E. (1991). Portfolioassessment in the reading-writing classroom. Nordwood, MA: Christopher Gordon.
- Topping, K. (1998). Peer-assessment between students in colleges and universities. *Review of Educational Research*, 68, 3, pp. 249-276.
- Vernon , D.T.A., & Blake, R.L. (1993). Does problem-based learning work? A meta-analysis of evaluative research. *Academic Medicine*, p. 68, pp. 550-563.
- Vygotski, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wilbrink, B. (1997). Assessment in historical perspective. *Studies in Educational Evaluation*, 22 (1), pp. 31-48.
- Williams, E. (1992). Student Attitudes towards Approaches to Learning and Assessment. *Assessment and Evaluation in Higher Education*, 17(1), pp. 45-58.
- Wolf Palmer, D. (1989). Portfolio assessment: Sampling student work. *Educational Leadership*, 46, pp.

35-39.

- Wolf, D., Bixby, J., Glenn III, J., & Gardner, (1991). To use their minds well: Investigating new forms of student assessment. *Review of Research in Education*, 17, pp. 31-73.
- Zoller, Z. & Ben-Chaim, D. (1997). Student self-assessment in HOCS Science Examinations: is it compatible with that of teachers? Paper presented at the meeting of the European Association for Research on Learning and Instruction, Greece, Athens, August 26-30.