

TECNOLOGÍAS SEMÁNTICAS PARA LA EVALUACIÓN EN RED: ANÁLISIS DE UNA EXPERIENCIA CON LA HERRAMIENTA OeLE

María del Mar Sánchez Vera⁽¹⁾, María Paz Prendes Espinosa⁽¹⁾, Jesualdo Tomás Fernández Breis⁽²⁾

⁽¹⁾ Departamento de Didáctica y Organización Escolar
Universidad de Murcia

⁽²⁾ Departamento de Informática y Sistemas
Universidad de Murcia

RESUMEN

No cuestionamos ya el importante papel que tienen las Tecnologías de la Información y las Comunicaciones (TIC) en nuestra realidad universitaria actual. El rápido devenir de estas herramientas ha supuesto la configuración de nuevos espacios de enseñanza-aprendizaje dentro de las distintas modalidades de enseñanza universitaria actuales. En este artículo se presentan los resultados de una investigación realizada tras la incorporación de un programa que permite realizar y corregir pruebas de evaluación de desarrollo a través de la red. Este programa permite hacer exámenes con preguntas de desarrollo y ofrece feedback al alumnado respecto a la prueba de evaluación realizada. Esta experiencia se llevó a cabo con un grupo de alumnos de la Licenciatura en Pedagogía de la Universidad de Murcia (España). Estos alumnos cursaban una asignatura optativa en red ofrecida por la titulación y se planteó la experiencia creando un entorno de evaluación en red que va más allá de las tradicionales pruebas tipo test utilizadas en los exámenes on-line. La finalidad de este análisis permite conocer no únicamente las posibilidades pedagógicas del entorno de evaluación en red, sino establecer además pautas de actuación para configurar futuros escenarios de evaluación en entornos virtuales con estas herramientas. Hemos de hacer constar que esta investigación ha sido realizada en el marco

Correspondencia:

María Paz Prendes Espinosa, Departamento de Didáctica y Organización Escolar. Facultad de Educación. Universidad de Murcia. E-mail: pazprend@um.es

del Proyecto 08756/PI/08 "Plataforma Semántica de Formación a la Carta" financiado por la Fundación Séneca (Comunidad Autónoma de la Región de Murcia).

Descriptores: Evaluación; Enseñanza superior; E-learning; Web Semántica.

SEMANTIC TECHNOLOGIES IN ONLINE ASSESSMENT: ANALYSIS OF THE USE OF OeLE

ABSTRACT

The key role of Information and Communication Technology (ICT) in universities is out of question. The fast development of ICTs has brought with it the creation of new teaching-learning environments in higher education. This paper shows the results of a study that made use of a software in order to conduct and assess short-answer online tests. The software also provided students with feedback on their performance. The study was carried out with a group of undergraduate Education students at the University of Murcia (Spain) enrolled in an online subject. The focus of the online assessment procedures introduced in this study was beyond multiple-choice tests traditionally used in online assessment. Results of this study not only showed the scope of online assessment, but also defined guidelines for evaluation in online courses. This research study is part of the "Semantic Environment for Personalized Learning" Project funded by the Fundación Séneca (Murcia, Spain).

Keywords: Assessment; Higher Education; E-learning; Semantic Web.

I. INTRODUCCIÓN: REVISIÓN TEÓRICA

La evaluación es un componente fundamental en el proceso de enseñanza-aprendizaje y, además, es un concepto amplio y multidimensional. De hecho, desde hace varios años es un tema de constante preocupación entre los docentes, las autoridades escolares y los investigadores de la educación (Perrenoud, 1990). Este tema sigue candente en la actualidad por las nuevas cuestiones pedagógicas que supone la implementación de las TIC. Al respecto, Pnuie, Zinnbauer y Cabrera (2006) explican que se puede identificar en la Unión Europea un impulso claro para intentar establecer una estrategia de evaluación *on-line* en la Educación Superior. Esta estrategia pasaría por tener en cuenta diversos componentes como el soporte que pueden dar las Tecnologías de la Información y la Comunicación (TIC) a sistemas de evaluación tales como el tipo test, y la completa implementación de sistemas de evaluación basados en TIC para la evaluación por ejemplo de las capacidades tecnológicas.

Además, en el reto que supone el Espacio Europeo de Educación Superior (EEES), las TIC tienen un papel fundamental no sólo como medio de comunicación con nuestro alumnado, sino como un entorno que puede facilitar experiencias semipresenciales o totalmente virtuales. Esto significa que hay que plantearse cómo se van a evaluar esas competencias que se tienen que trabajar con los alumnos y al mismo tiempo plantearse nuevos sistemas de evaluación basados en TIC con todas las connotaciones pedagógicas que ello conlleva.

La realidad de la enseñanza apoyada en TIC nos lleva a hablar de procesos de enseñanza-aprendizaje presenciales, semipresenciales (*blendedlearning*) o virtuales (*eLearning*). Si consideramos al mismo tiempo la evaluación educativa como un componente esencial de los procesos educativos, podemos preguntarnos entonces qué pueden aportarnos las TIC en experiencias de evaluación en red. En este sentido, Barberá (2006:7) identifica tres aportaciones que han hecho las TIC a la evaluación educativa. La primera de ellas es permitir una *evaluación automática*. La principal ventaja de la evaluación automática se refiere al hecho de posibilitar una respuesta rápida. En muchos casos el alumno puede conocer en el momento si su respuesta es errónea o no, lo que supone un *feedback* inmediato con las ventajas que esto conlleva, y que permite también la autoevaluación. El inconveniente de la evaluación automática radica en que la comunicación entre el docente y el estudiante es limitada ya que, al realizarse normalmente mediante pruebas estandarizadas, se omite la personalización de las respuestas.

La segunda aportación que han hecho las TIC a la evaluación nos permite hablar de una *evaluación enciclopédica*. Para el alumno este proceso es rápido y cómodo debido a la gran cantidad de información digitalizada que existe, por lo que la elaboración de documentos puede ser más sencilla (teniendo en cuenta el riesgo de que el alumno realice un "corta-pegar", a pesar de que podemos encontrar algunos programas para detectar este hecho). Este tipo de evaluaciones hasta el momento se han limitado a pruebas objetivas (Prendes, 2007a), algo que podría cambiar con nuevas herramientas como puede ser la Web Semántica.

Finalmente podemos indicar una tercera aportación de las TIC a la evaluación, que se refiere a la *evaluación colaborativa*. A pesar de la gran cantidad de entornos de aprendizaje colaborativo en red (CSCL, BSCW, etc.) la evaluación colaborativa no es muy utilizada en los cursos de *eLearning*, quizás porque no se saben identificar claramente los indicadores para la evaluación de los aprendizajes. Sin embargo, se deberían aprovechar las ventajas que nos ofrecen las TIC de evaluar no sólo el producto, sino el proceso de aprendizaje de los alumnos.

Como vemos, las TIC pueden aportar muchas alternativas a la evaluación, pero a pesar de ello, como indica Prendes (2007b:218), "la evaluación en tele-enseñanza es uno de los temas que quizá, a pesar de ser objeto de estudio habitual, menos cambios ha sufrido". Hablar de la incorporación de las TIC en procesos de enseñanza-aprendizaje supone hablar de nuevos roles de profesor y alumno, así como de una transformación en las estrategias de enseñanza y aprendizaje. Sin embargo, se suele observar una tendencia en la evaluación universitaria en red hacia modelos tradicionales (como exámenes vigilados, tareas escritas, etc.). Esta opinión la comparten otros autores (Driscoll, 2001; Strither, 2002; Castillo, 2002; Rubio, 2003; De Pablos, 2004; Prendes, 2007a; Steban Git *et al*, 2009; Radenkovic, Krdzavac y Devedzic, 2010, Castellanos Nieves, 2010) que vienen a exponernos la problemática que encontramos respecto a cómo evaluar a nuestros alumnos en red, que se plasma en que la evaluación en *eLearning* se suele realizar a través de una prueba estandarizada (normalmente tipo test) a través del ordenador o de un examen presencial incluso aunque el curso sea totalmente a distancia y virtual, debido seguramente a la necesidad de acreditación de los estudiantes en la Educación Superior. Además, estos autores nos destacan otra problemática importante y es que la retroalimentación (*feedback*) en la evaluación en *eLearning* es un elemento importante

que no se suele tener en cuenta en los mecanismos que se realizan para evaluar a los alumnos. Hay falta de sistemas para la reutilización de contenidos formativos.

Para poder crear un sistema de evaluación en *eLearning* que tenga en cuenta estas peculiaridades hay que tener en cuenta que los ordenadores no entienden las palabras, los ordenadores procesan signos. El procesamiento del lenguaje natural se ocupa de la formulación e investigación de mecanismos eficaces computacionalmente para la comunicación entre personas o entre personas y máquinas por medio del lenguaje natural (Fernández Breis *et al* 2007).

La Web Semántica no es un elemento innovador en internet, el propio Berners Lee intentó incluir información semántica en su idea del WWW, pero no fue posible debido a las tecnologías disponibles por aquel entonces. Se retoma este concepto de Web Semántica ahora para referirnos al “conjunto de iniciativas, tecnológicas en su mayor parte, destinadas a crear una futura World Wide Web en la cual los ordenadores puedan procesar la información” (Berners-Lee *et al*, 2001). Una de las tecnologías fundamentales para el desarrollo de la Web Semántica es la ontología. Una ontología es esencialmente un modelo conceptual de información formal y estructurada que podemos ver como un modelo semántico que contiene conceptos y sus relaciones dentro de un dominio. Una de sus grandes ventajas es su formalización, por lo que pueden ser entendidas por las máquinas. Gracias al trabajo con ontologías, le indicamos al ordenador un esquema con sentido del contenido a tratar, de modo que posteriormente la tecnología puede navegar por una estructura organizada de información. En la práctica, uno de los motivos básicos por los que la ontología ha cobrado tanta importancia y se ha extendido su uso es por las ventajas que posee (Fernández-Breis y Martínez-Béjar, 2002): i) reusabilidad, una misma ontología se puede reutilizar en diversas aplicaciones de forma individual o combinada con otras y ii) compartición, el conocimiento que incluye permite que sea compartido por una determinada comunidad.

Si el hipertexto nos posibilitó la interactividad gracias a poder movernos de un enlace a otro, la Web Semántica intenta mejorar los mecanismos para compartir información y recursos. Las tecnologías de la Web Semántica facilitan la gestión del conocimiento y promueven la interoperabilidad, imponiendo un entorno de trabajo en el que cada sistema emplea el significado de los datos en diferentes contextos. La Web Semántica pretende convertirse en un mediador universal para el intercambio de información. Por otro lado, la Web Semántica proporciona mecanismos para la creación de recursos semánticos con el conocimiento común de un dominio. Estos recursos son utilizados por personas y aplicaciones que necesitan compartir información sobre determinado dominio, entendiendo por dominio la especificación de un área de conocimiento. Aquí es dónde las ontologías ejercen su papel protagonista como soporte del conocimiento semántico.

Acercándonos a ámbitos educativos, no podemos dejar pasar la oportunidad de participar en estas nuevas realidades tecnológicas. La Web Semántica puede proveer al *eLearning* de materiales de aprendizaje compartibles e intercambiables, lo que permite la personalización de los contenidos. Al mismo tiempo, la Web Semántica puede ser tratada como una plataforma muy aconsejable para implementar un sistema de *eLearning*, al proporcionar mecanismos para el desarrollo de ontologías para el aprendizaje,

la anotación semántica de materiales, su combinación para la definición de cursos y su posterior evaluación (Castellanos Nieves *et al*, 2011).

En este marco surge la colaboración multidisciplinar del *Grupo de Tecnologías de Modelado, Gestión y Procesamiento de Conocimiento* (Facultad de Informática) y el *Grupo de Investigación de Tecnología Educativa* (Facultad de Educación), ambos de la Universidad de Murcia, para analizar las posibilidades de una herramienta telemática creada y patentada por dicho grupo, la plataforma OeLE (Ontology eLearning Environment). Esta plataforma permite realizar exámenes *on-line* mediante preguntas abiertas y que los alumnos reciban un *feedback* de su proceso de aprendizaje.

2. PROPÓSITO DE LA INVESTIGACIÓN

2.1. OeLE: una herramienta para la evaluación en red

OeLE es una herramienta que usa tecnologías de Web Semántica y permite realizar pruebas escritas que van más allá de un examen final tipo test que tradicionalmente se utiliza en los exámenes *on-line*. Una de las grandes posibilidades que nos ofrece además es, a partir de los conceptos que el alumno desarrolla, ofrecerle información acerca de su prueba indicando lo que ha hecho o no bien, es decir, proporcionarle un *feedback* al alumno en su proceso de aprendizaje.

Las ontologías son el motor de la herramienta OeLE. Esta herramienta hace uso de distintos tipos de ontologías. Por un lado, el conocimiento de cada curso se expresa mediante una ontología. Por otro lado, una ontología de anotaciones facilita un modelo fácilmente extensible, compartible y reutilizable en caso de que surjan nuevos requisitos, nuevos tipos de anotaciones y pueda ser empleado de manera eficiente en otras aplicaciones. OeLE procesa ontologías implementadas en el lenguaje OWL, que es el estándar del W3C para el intercambio de contenidos semánticos. Las anotaciones que se realizan para las preguntas y respuestas que componen las pruebas de evaluación son instancias del correspondiente tipo de anotación contenido en este modelo.

A las preguntas se les asocian unas anotaciones, las cuales conllevan una puntuación cuantitativa. Estas anotaciones se refieren a la información con que el alumno debe responder en cada pregunta. Cada respuesta del alumno tiene asociada una anotación formada por una lista de los mismos tipos de elementos ontológicos que en el caso anterior, aunque sin llevar asociado un valor de puntuación. Dichas anotaciones también están relacionadas con el conocimiento del curso. Lo que hace OeLE es analizar la similitud entre los elementos de respuesta del alumno y los de la respuesta esperada (correcta). Para ello, en primer lugar se analiza la similitud entre elementos ontológicos del mismo tipo. Para medir la similitud existente entre dos elementos ontológicos del mismo tipo se utiliza la composición de una serie de métricas que nos permiten compararlos de acuerdo a ciertas características estructurales y lingüísticas, cuyo resultado estará entre 0 para una total diferencia y 1 para una total igualdad.

OeLE es una herramienta creada a partir de software libre y es un entorno multiplataforma, es decir, puede utilizarse en cualquier ordenador sin tener que adquirir licencia para ello y funciona en todos los ordenadores independientemente del Sistema Operativo utilizado. La aplicación para profesores y administradores

está desarrollada en Java mientras que la aplicación para los estudiantes, para conseguir un mayor nivel de flexibilidad de acceso a los exámenes, requiere de forma exclusiva de un navegador. De este modo el estudiante no necesita instalar nada en su ordenador y le basta con acceder al servicio Web a través de cualquiera de los navegadores disponibles.

2.2. Entorno de evaluación

Esta experiencia se desarrolla en el marco de la asignatura “Diseño y Evaluación de Materiales Didácticos” (área de Didáctica y Organización Escolar), es una optativa de 2º ciclo con 4,5 créditos de libre configuración y se impartió en el segundo cuatrimestre del curso 2008-2009 en la Licenciatura en Pedagogía de la Universidad de Murcia.

La metodología de trabajo propuesta en la asignatura contempla un modelo de eLearning basado en la individualización de la enseñanza y en el trabajo autónomo del alumno a través del Campus Virtual (que en el caso que nos ocupa se desarrolló en Suma Docente en su versión nueva, SUMA v2.0).

La evaluación planteada en la asignatura se centró en el proceso y en los resultados de aprendizaje, es decir, se consideró tanto la evaluación de procesos como de productos. Como bien afirma Martín (2007) la evaluación de procesos permite al alumno poner en práctica lo que van aprendiendo, lo que nos proporciona información que debemos tener en cuenta en la evaluación.

Los criterios definidos dentro de la evaluación formativa fueron los siguientes:

- El grado de participación individual en las actividades previstas (chats, foros y videoconferencias).
- El grado de utilización de la tutoría (interacción con el profesor).
- El grado de interacción con el grupo, en el marco del trabajo colaborativo (rincón del alumno).
- La cumplimentación del blog personal con carácter semanal (portafolio).

Los criterios que se tuvieron en cuenta en la evaluación sumativa fueron:

- La realización de las actividades indicadas para cada tema.
- La entrega del trabajo individual (portafolio).
- La realización de un examen final en un entorno de evaluación y la consulta del *feedback*.

Es en este último punto en el que el entorno de evaluación *on-line* se utilizó y en donde se enmarca esta investigación. Las demás variables que forman parte de la asignatura se tienen en cuenta a la hora de diseñar, analizar y resumir la experiencia, sin embargo, el foco de atención estaba en comprobar la utilidad educativa del entorno de evaluación innovador compuesto por herramientas de Web Semántica y sistemas de *feedback*.

Retomando nuestra investigación, el escenario en el que ésta se desenvuelve es el entorno de evaluación en red creado en el contexto de la asignatura:

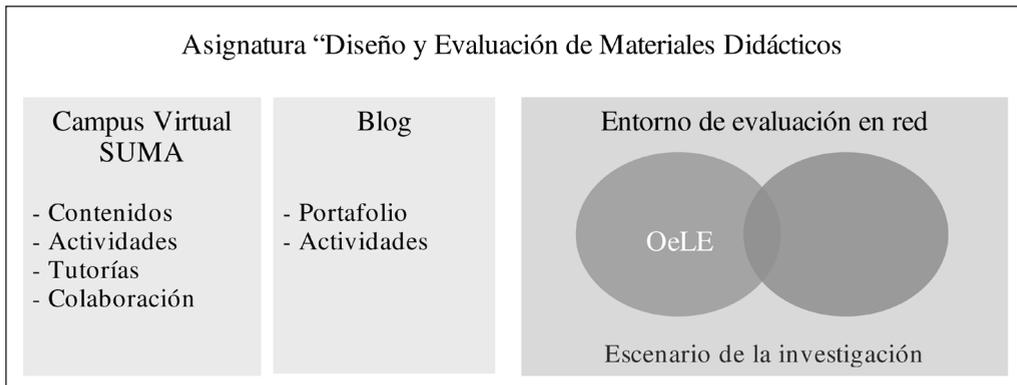


FIGURA 1
ESCENARIO DE LA INVESTIGACIÓN UBICADO EN EL DESARROLLO DEL CURSO VIRTUAL

Como vemos, esta investigación se ha centrado principalmente en el entorno de evaluación de la asignatura, el cual se compone principalmente de dos herramientas: OeLE (programa para la evaluación *on-line* con Web Semántica) y la página Web de contenidos de refuerzo.

En la asignatura de “Diseño y Evaluación de Materiales Educativos” la experiencia en el entorno de evaluación se desarrolló al final de la materia (evaluación sumativa), sin embargo es importante indicar que la calificación final del alumno no dependió sólo de esta prueba sino también de otros elementos evaluativos (básicamente actividades y portafolio).

2.3. Objetivos

Nuestro primer objetivo fue conocer la validez del programa, es decir, si la corrección que hacemos con la aplicación OeLE sería similar a la corrección realizada manualmente, sin necesidad del programa. Una vez validado pretendemos conocer cómo funciona el nuevo entorno de evaluación *on-line* que proporciona la Web Semántica a partir de las valoraciones que realizan los alumnos de la experiencia. De acuerdo con Cabero (1998) los nuevos medios nos ofrecen la posibilidad de crear entornos más flexibles y abiertos para el aprendizaje, los cuales facilitan al alumno la adquisición de aprendizajes significativos. Anteriormente hemos indicado las potenciales ventajas que ofrece la Web Semántica. La idea inicial de la investigación fue promover un entorno novedoso de evaluación basado en estas tecnologías de Web Semántica que permiten, entre otras cosas, realizar exámenes *on-line* de preguntas abiertas, y de este modo comprobar su efectividad didáctica. Como hemos indicado, el programa permite la retroalimentación e informa a los alumnos de los fallos que han cometido, por lo que el segundo gran objetivo fue el de valorar la utilidad del *feedback* desde el punto de vista de los alumnos.

Es importante destacar que todo el proceso se desarrolló de forma virtual, tanto la realización de los exámenes como el cuestionario. Los aspectos sometidos al análisis

sis fueron los conocimientos teóricos que el alumno había adquirido a lo largo de la asignatura y que fueron preguntados en el entorno de evaluación en red a través de exámenes de desarrollo, exámenes que se realizaron en dos momentos: antes y después de un proceso intermedio de *feedback*. De manera simplificada, podemos ver cómo ha funcionado el entorno de evaluación en la siguiente figura:

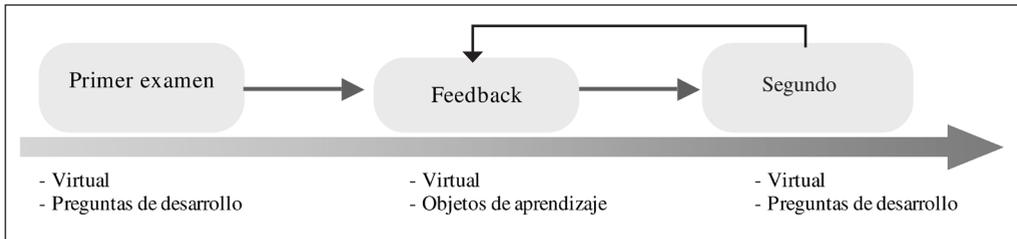


FIGURA 2
PROCESO REALIZADO POR LOS ALUMNOS EN EL ENTORNO DE EVALUACIÓN

2.4. Metodología de la investigación

De acuerdo con las premisas de Ramos, Suárez y Jornett (2006) hay que considerar que la validación de cualquier sistema de evaluación se basa en una acumulación de evidencias acerca de la adecuación del funcionamiento del mismo. Es decir, un modelo no se valida para todos sus usos posibles, sino que a partir de unas circunstancias concretas podemos conocer los resultados que ha ofrecido a partir de un programa de formación concreto.

El proceso de trabajo que se realizó fue el siguiente: los alumnos realizaron un primer examen de preguntas abiertas sobre los contenidos de la asignatura, posteriormente se les ofreció un *feedback* de sus resultados a través de la herramienta OeLE. Los propios estudiantes podían acceder al programa y conocer en cada pregunta cuáles eran los conceptos que habían respondido bien, mal o que necesitaban mejorar. Durante una semana los estudiantes podían dirigirse a una página web para consultar los fallos o mejorar su conocimiento respecto al tema. La página Web para los contenidos de refuerzo a consultar como consecuencia del proceso de *feedback* se realizó con el programa *ExeLearning* y se estructuró en función de las preguntas del examen, en las que se organizaba la información mediante Objetos de Aprendizaje. Esta página Web se puede observar en <http://miuras.inf.um.es/~oele/objetos/>. Posteriormente a la consulta de los fallos, los alumnos realizaron un segundo examen sobre el mismo contenido.

El muestreo realizado es no aleatorio y abarcó a todos los alumnos que estaban matriculados y se presentaron al examen de la asignatura. En este caso, la *muestra invitada* fue de 26 alumnos, que representa el total de estudiantes que estaban matriculados de la asignatura virtual. La *muestra participante* fue de 22 alumnos que coincide con la *muestra real* de los estudiantes de los que poseemos datos estadísticos, lo cual representa también la muestra productora, la cual se corresponde con el número de individuos de los que se ha obtenido información en cualquiera de las fases de investigación.

El método utilizado es experimental. Son varios los autores (Tejedor, 1984; Arnal, Del Rincón y Latorre, 1992; Corbetta, 2007) que indican que en el experimento en Ciencias Sociales la finalidad del experimentador es que los cambios o variaciones en los valores de la variable dependiente sean atribuidos a las manipulaciones de la variable independiente. De hecho, siguiendo las ideas de Hernández Pina (2001), el método experimental es el que mejor garantiza las relaciones causales. El término de diseño experimental hace referencia al plan de asignación de los sujetos a los grupos, las condiciones experimentales a las que son sometidos los sujetos y al análisis estadístico asociado con el plan.

Respecto a los instrumentos de recogida de información es importante indicar que, en esta investigación, la propia experiencia en sí supone la realización de una prueba de evaluación, lo cual ya de por sí nos va a permitir acceder a una serie de datos que son las calificaciones de los alumnos en tres sentidos: la calificación automática que realiza el programa OeLE del primer examen, la calificación manual que realiza el profesor del primer examen para poder validar la herramienta, y finalmente, la calificación del segundo examen tras la experiencia de *feedback*.

Sin embargo, estos datos no son suficientes. Se puede afirmar sin aventurarnos demasiado que un alumno que realiza un segundo examen tras haber recibido un *feedback* (y teniendo en cuenta que el segundo examen versa sobre los mismos contenidos que el primero) va a obtener una calificación superior o al menos parecida a la del primer examen, por tanto, no podemos valorar la utilidad del *feedback* únicamente a través de ese procedimiento, por ello se utilizó también el cuestionario final.

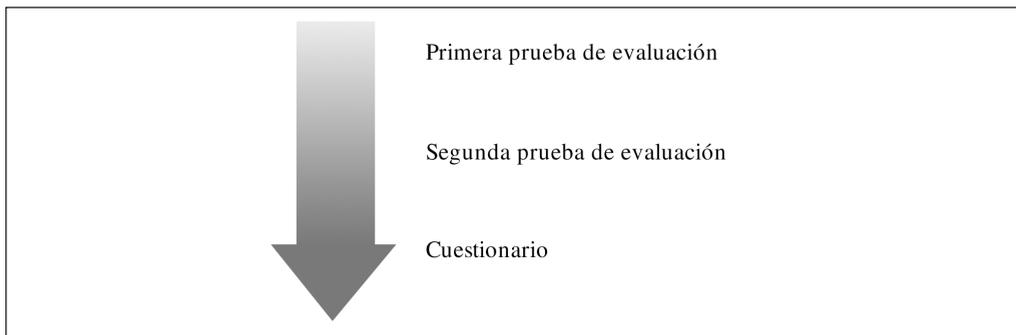


FIGURA 3
INSTRUMENTOS UTILIZADOS

A la hora de plantearse la construcción del cuestionario se tuvieron en cuenta los aspectos referidos a los métodos de escalamiento. El cuestionario está compuesto por 28 preguntas tipo Likert, pidiendo a los encuestados que manifiesten su grado de acuerdo o desacuerdo con el ítem propuesto. Las opciones de respuesta han sido: muy en desacuerdo, en desacuerdo, de acuerdo, muy de acuerdo. A pesar de que las primeras escalas tipo Likert comprendían siete alternativas (Corbetta, 2007), hoy en día se han reducido a cinco opciones de respuesta, incluso a cuatro, como es el caso del presente cuestionario, que supone la eliminación de la categoría intermedia. Es cierto

que los aspectos de deseabilidad social suelen acentuarse en cuestionarios referidos a personalidad, sin embargo, consideramos que era interesante reducir las alternativas de la escala Likert a cuatro ítems de modo que los encuestados tuvieran que decantarse por un aspecto en concreto. En cada uno de los ítems se le añadió también la opción de que el alumno respondiera “No sabe/No contesta”. Como afirma Corbetta (2007) es importante que se incluyan las respuestas “no sé”, y que no todas las afirmaciones tengan el mismo sentido, a fin de obligar a los entrevistados a reflexionar para que sus respuestas sean coherentes.

El cuestionario fue validado mediante el juicio de expertos. Concretamente se especificó que era necesario diferenciar entre tres aspectos: valoración de los aspectos técnicos de OeLE, valoración de los aspectos técnicos de la Web con OA para el *feedback* y valoración de los aspectos referidos a la metodología de evaluación *on-line* empleada (es decir, la valoración de cómo había funcionado todo el entorno de evaluación). Así mismo, el juicio de los expertos nos llevó a reformular alguna de las cuestiones referidas a la última parte del cuestionario, la cuál trata de conocer qué aspectos fueron los que más consultaron los estudiantes en el *feedback*.

La fiabilidad del cuestionario ha sido realizada mediante el coeficiente *alfa de Cronbach*, el cual es posible extraer en el SPSS mediante la opción de realizar un análisis de fiabilidad. La obtención del estadístico de fiabilidad *Alfa de Cronbach* nos ofrece una puntuación de 0,750, lo que indica una fiabilidad alta (Díaz de la Rada, 1999). Esto nos lleva a considerar como aceptables los índices de fiabilidad del cuestionario.

Retomando el desarrollo de la experiencia, hay que indicar que los alumnos hicieron el examen a través de OeLE y la misma aplicación OeLE nos proporciona datos, teniendo en cuenta que precisamente, uno de los focos de interés de la investigación es conocer su funcionamiento y saber su utilidad. Una vez los alumnos hicieron la prueba de evaluación correspondiente, para poder concretar las calificaciones es necesario realizar un procedimiento de corrección de las pruebas en la plataforma. OeLE permite acceder a la respuesta de cada uno de los alumnos. En esta ventana podemos visualizar la ontología de la respuesta esperada (la respuesta “ideal”) del curso e ir anotando los elementos que identificamos en la respuesta del alumno, de este modo posteriormente se corrigen las preguntas y se obtiene la calificación del alumno. La calificación automática es un aspecto que se encuentra en proceso de desarrollo del grupo de investigación, con vistas a aumentar el potencial de la aplicación cuando este proceso se haga mecánicamente.

Por tanto, las calificaciones del examen las proporciona OeLE, que automáticamente crea el *feedback* para los alumnos y los docentes. Posteriormente una de las profesoras de la asignatura realizó la corrección manual (sin tener información previa sobre la corrección realizada por OeLE) y también contamos con la calificación que nos ofrece OeLE del segundo examen.

Para conocer si se podían cumplir los objetivos planteados, en varios casos ha sido necesario utilizar técnicas de estadística confirmatoria para saber si existían diferencias estadísticamente significativas a un nivel de significación de 0,05 en los diferentes momentos de la investigación. Para ello se ha utilizado el programa de análisis estadístico SPSS.

Respuesta a Pregunta Abierta - Corregir

Datos de la pregunta

Enunciado
Identifica y define qué recurso sonoro surge a partir de un término

Respuesta esperada
Podcast. El origen de la palabra Podcast surge de la unión de las palabras ipod y broadcasting. Un Ipod es un reproductor portátil comercializado por Apple, y broadcast se refiere a la emisión de la televisión o la radio. El término podcast aparece en el 2004, es acuñado por Ben Hammersley en The Guardian para describir la posibilidad de escuchar la radio en los reproductores portátiles.

Anotaciones de la respuesta esperada
(Muy alto, 5) Concepto Adquisición_sencilla
(Bajo, 2) Concepto Apple
(Bajo, 2) Concepto Ben_Hammersley
(Alto, 4) Concepto Broadcasting
(Bajo, 2) Concepto Característica
(Muy alto, 5) Concepto Claridad
(Muy alto, 5) Concepto Elegir_contenido
(Bajo, 2) Concepto Fabricante

Puntuación máxima 2

Datos de la respuesta

Respuesta obtenida
El recurso sonoro que surge es el Podcast, término que aparece en 2004 acuñado por Ben Hammersley. El término podcasting en el período se utilizó por primera vez el 12 de febrero de 2004 en el periódico The Guardian, cuando hablamos de podcasting estamos hablando de la creación de archivos de sonido o de vídeo, para su distribución es necesario que el usuario se descargue de Internet un

Anotaciones de la respuesta obtenida
("para adquirirlo...") Concepto Adquisición_sencilla
("acuñado por ben hammersley") Concepto Ben_Hammersley
("ventajas que ofrece") Concepto Característica
("elección del tema") Concepto Elegir_contenido
("normalmente es gratuito") Concepto Gratuito
("Ipod") Concepto Ipod
("se necesita un programa de sindicación") Concepto Itunes
("Bueno") Concepto Originalidad

Aspectos a mejorar
(Bajo, 2) Concepto Apple
(Alto, 4) Concepto Broadcasting
(Muy alto, 5) Concepto Claridad
(Bajo, 2) Concepto Fabricante

Datos de la calificación

Calificación Manual 1,50 Calificación Automática 1,14

Corregir Automáticamente

Modificar Recargar

FIGURA 4
PERSPECTIVA DEL PROFESOR. CALIFICACIÓN DE UNA PREGUNTA ABIERTA

En primer lugar encontramos el análisis de las calificaciones de los alumnos, realizado mediante el análisis de varianzas con medidas repetidas, en el que se han incluido también las medias de las distintas pruebas. En los modelos de medidas repetidas es necesario suponer que las varianzas de las diferencias entre cada dos niveles de medida repetida son iguales. Esto quiere decir que en nuestros tres niveles tenemos tres pares de combinaciones: 1-2, 2-3, 1-3 (siendo 1 la calificación manual del primer examen, 2 la calificación que ofrece OeLE del primer examen y 3 la calificación que ofrece OeLE del segundo examen). Por tanto, para este análisis el factor sería el "método de corrección empleado", siendo los niveles del factor los tres valores mencionados.

Por otro lado, las técnicas estadísticas para el análisis de los datos del cuestionario, debido al número de sujetos de la muestra, se refieren a la estadística no paramétrica, ya que el contar con menos de 30 sujetos es un supuesto paramétrico importante para decidir utilizar técnicas de carácter paramétrico. En el caso anterior (referido al análisis de los exámenes) la justificación para realizar un ANOVA de medidas repetidas viene de la mano de la idea de que este tipo de análisis supone una gran potencia

estadística a la hora de fundamentar algún tipo de conclusión (Ato y Vallejo, 2007). Somos conscientes del tamaño muestral, sin embargo, tenemos en consideración que en el caso de los diseños de medidas repetidas, el número total de observaciones se multiplica por el número de medidas con las que contamos, y de este modo para cada individuo no se toma únicamente un registro de medida. Es decir, de los 22 casos con los que contamos, el diseño de medidas totalmente repetidas lo multiplica de tal modo que contamos con 66 observaciones, lo que supone un número mayor a 30 sujetos. Teniendo en cuenta también que el análisis de medidas repetidas produce resultados significativos se puede presumir de una potencia considerable para los resultados (Ato y Vallejo, 2007).

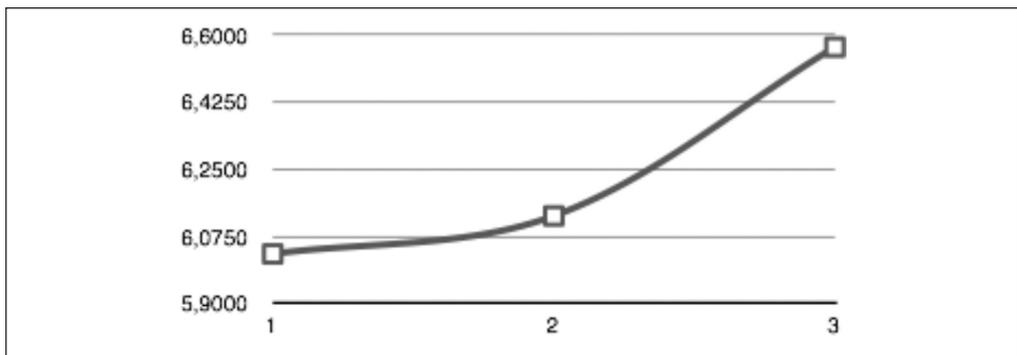
3. RESULTADOS Y CONCLUSIONES

Los primeros resultados que obtenemos son los que nos arroja el propio programa OeLE, que se refieren a la nota de cada alumno en función del método de corrección:

TABLA 1
EJEMPLO DE LAS CALIFICACIONES DE LOS ALUMNOS

	1 ^{er} examen corrección manual	1 ^{er} examen corrección OeLE	2 ^o examen corrección OeLE
Alumno 1	6,8	7,2	8,01
Alumno 2	5,65	5,79	6,9
Alumno 3	5,8	5,83	6,39
Alumno 4	5,6	6,12	6,58
...

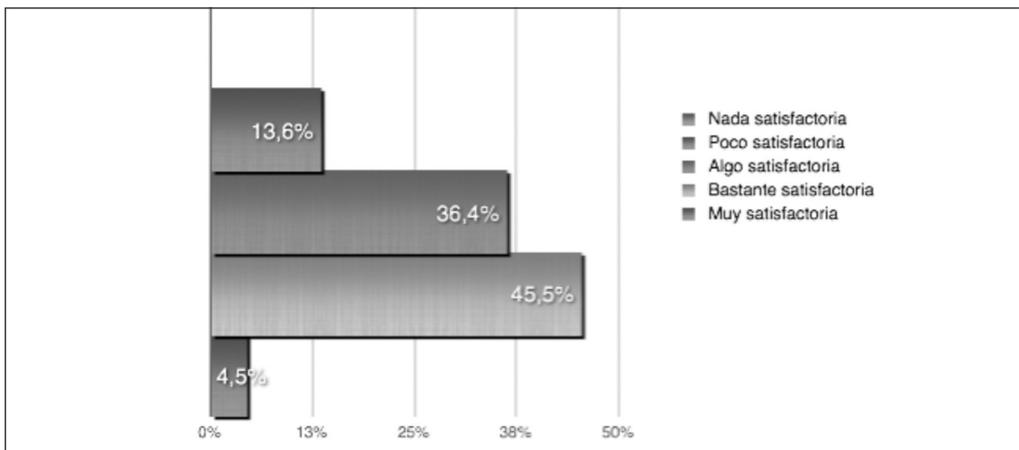
Extrayendo la media de las respuestas del alumno para cada calificación (siendo 1: nota media del primer examen, corrección manual; 2: nota media del primer examen, corrección OeLE; y 3: nota media del segundo examen, corrección OeLE), encontramos:



GRÁFICA 1
MEDIAS MARGINALES ESTIMADAS DEL TIPO DE CALIFICACIÓN

Como se puede comprobar, la media de las calificaciones para los tres factores nos indica que la media de la calificación del programa es un 0,099 mayor y que la media de calificaciones del segundo examen respecto al primero (correcciones ambas del programa OeLE) es mayor en un 0,539. Para conocer si estas diferencias son significativas o no se procedió a realizar distintas técnicas de análisis. El análisis estadístico de medidas repetidas nos muestra que no existen diferencias significativas entre la corrección manual y la corrección automática (la que realiza el programa).

Adentrándonos en los resultados que nos arroja el cuestionario que han respondido los alumnos es importante indicar que la cuestión referida al nivel de satisfacción con la experiencia realizada ha ofrecido el siguiente dato: más de la mitad de los alumnos señalan su nivel de satisfacción con la experiencia como bastante o muy satisfactoria (gráfica 2). En concreto, el 13,6% consideró que ha sido poco satisfactoria, el 36,4% indicó que ha sido algo satisfactoria, el 45,5% que ha sido bastante satisfactoria, y el 4,5% que ha sido muy satisfactoria. Ningún alumno marcó que la experiencia hubiera sido nada satisfactoria. De hecho para el grado de satisfacción encontramos una media de 3,41 (desviación típica de 0,796). Esta media se encuentra por encima del nivel medio 3 que es “algo satisfactorio”.



GRÁFICA 2

INDICA EL GRADO DE SATISFACCIÓN CON LA METODOLOGÍA DE EVALUACIÓN DE PREGUNTAS ABIERTAS EMPLEADA

El por qué de ese grado de satisfacción es contestado por el alumno mediante una pregunta abierta en el cuestionario, en el que encontramos diversos tipos de respuesta. Los estudiantes que se muestran muy satisfechos hacen inciso en que han considerado la experiencia interesante porque este tipo de examen de preguntas abiertas les ha permitido reflexionar y expresarse mejor, también han destacado la novedad que les supone realizar un examen de desarrollo en Internet. Los estudiantes que se muestran algo satisfechos indican aspectos positivos como que les ha parecido interesante la experiencia pero que el tiempo para hacer el examen no fue suficiente. Los alumnos que indican que están poco satisfechos señalan que eso es debido a que este tipo de

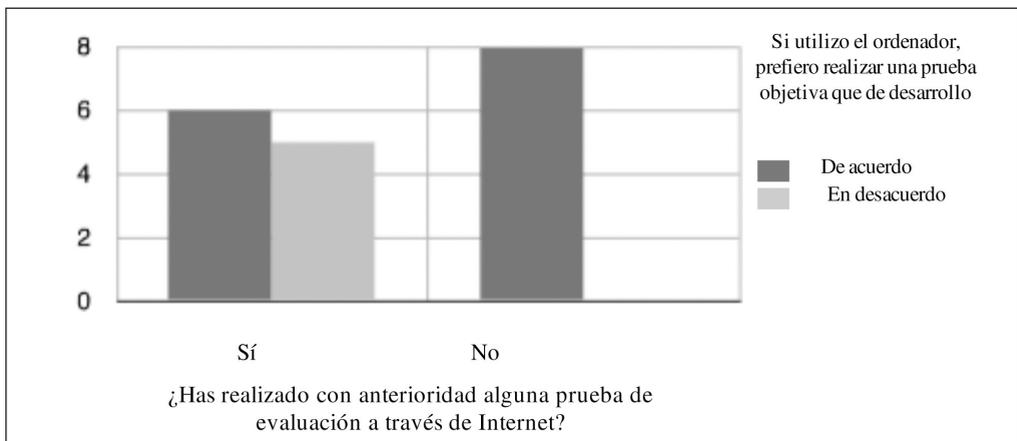
examen favorece el corta-pegar y que prefieren realizar un examen tipo test que de desarrollo, porque los exámenes de desarrollo son más subjetivos.

El análisis de frecuencias nos ha mostrado que, a pesar de que la mayoría de los alumnos no había participado con anterioridad en una experiencia de formación en red, un 59,1% indicó que sí había realizado con anterioridad alguna prueba de evaluación a través de Internet.

En consonancia con estos resultados, uno de los aspectos que consideramos importante era conocer si existe una relación significativa entre la experiencia previa y la preferencia respecto al tipo de examen que prefieren realizar se ha hecho; el cruce de variables nominales con una tabla de contingencia nos ha permitido observar que existe relación significativa ($p=0,026$) entre si se ha realizado una prueba de evaluación con anterioridad a través de Internet y la preferencia respecto a realizar una prueba tipo test en entornos de evaluación *on-line*. La magnitud de la asociación dada por la *V de Cramer* es de 0,510 lo que supone una asociación moderada. El *coeficiente de incertidumbre* que obtenemos es de 0,308, lo que nos indica que reducimos en un 30,8% el error que cometeríamos al pronosticar la variable preferencia del tipo test cuando tenemos en cuenta si el alumno ha realizado con anterioridad un examen o no.

Además comprobamos que esta preferencia hacia el tipo test está también relacionada con el grado de satisfacción mostrado de manera que cuando aumenta el grado de satisfacción con la experiencia encontramos disminuye la preferencia por el tipo test en exámenes en red. Hay que tener en cuenta que la experiencia previa del alumnado en más de un 80% se relacionaba con experiencias de exámenes tipo test.

Esta forma de evaluación en red a través de preguntas abiertas que el alumno tiene que desarrollar es diferente a la manera tradicional de desarrollar exámenes en un entorno presencial (con lápiz y papel). Estaríamos hablando en este caso de que el examen puede ser el mismo (un examen de desarrollo con preguntas abiertas) pero que



GRÁFICA 3

RECUEENTO DE LA TABLA DE CONTINGENCIA PARA "SI UTILIZO EL ORDENADOR PREFIERO REALIZAR UNA PRUEBA OBJETIVA QUE DE DESARROLLO" Y "¿HAS REALIZADO CON ANTERIORIDAD ALGUNA PRUEBA DE EVALUACIÓN A TRAVÉS DE INTERNET?"

el entorno varía de la presencialidad a la virtualidad. A este respecto encontramos que los alumnos indican que el hecho de que el examen sea a través de internet les supone una metodología de estudio diferente a si fuera un examen presencial. Encontramos un 54,5% que está muy de acuerdo con la afirmación y un 27,3% que está de acuerdo, que en total representa el 81,8%.

Además, a la mayoría de estudiantes le ha gustado realizar un examen de desarrollo mediante la red (72,8% están de acuerdo o muy de acuerdo), aunque encontramos un 22,7% que están en desacuerdo y un 4,5% que están muy en desacuerdo con esta afirmación. Los aspectos técnicos del programa (color, tipo de letra, facilidad de uso) son valorados también de manera positiva por el alumnado participante de la experiencia.

Respecto a la percepción del *feedback* como elemento integrante del sistema de evaluación, entre otras cosas, se les preguntó acerca de si creían que la consulta de los fallos era algo innecesario, una pérdida de tiempo. A este respecto, un gran porcentaje de los encuestados encontraron que la consulta de los fallos no era una pérdida de tiempo (86,4%). Además el 77,3% afirmó que el material de *feedback* les sirvió para entender mejor algunos contenidos de la asignatura.

4. DISCUSIÓN

Lo primero que debemos indicar es que somos conscientes de que la incorporación de un medio, por sí mismo, no va a hacer que los alumnos aprendan mejor o que se favorezca un entorno de evaluación más completo por sí mismo. Pero aún con todas sus limitaciones, las herramientas de Web Semántica, en este caso OeLE, pueden convertirse en una importante ayuda para la evaluación en red, como ya lo son actualmente en otros ámbitos tales como el acceso a los contenidos o la comunicación entre los agentes educativos.

Realizar una prueba de evaluación de preguntas abiertas en red ha sido valorado como algo positivo por parte de los alumnos, tal y como se evidencia en el análisis estadístico que hemos realizado. En todos los casos, más de la mitad de los alumnos consideran que les ha gustado realizar un examen de desarrollo mediante la red y además la experiencia en general ha sido valorada con una media elevada.

Sin embargo, hay un punto en el que debemos incidir. Es curioso o cuanto menos interesante que los alumnos indiquen que les ha gustado la experiencia, pero luego indiquen que prefieren realizar un examen tipo test que de desarrollo en entornos en red. Nuestra hipótesis de partida nos permitió cruzar una serie de variables que nos ofrecieron unos resultados bastante interesantes en este aspecto. Encontramos que la experiencia previa de los alumnos (el hecho de haber realizado con anterioridad un examen en red) tiene relación con la respuesta dada respecto a qué tipo de prueba prefieren. Teniendo en cuenta que el análisis de las respuestas nos ha llevado a conocer que la mayoría de los alumnos que tenían experiencia previa realizando exámenes en red se refería a exámenes tipo test, podemos pensar que el hecho de haber realizado un examen tipo test con anterioridad influye en que sea este tipo de exámenes precisamente los que ellos prefieren realizar. Podemos enfocarlo desde la perspectiva de que un examen tipo test les sea más sencillo porque es a lo que ellos están acostumbrados, sin embargo, que prefieran el tipo test no implica que no les parezca adecuada la meto-

dología de exámenes de preguntas abiertas, que de hecho es bastante bien valorada por ellos mismos. Los alumnos que indican que están poco satisfechos señalan que eso es debido a que prefieren realizar un examen tipo test a uno de desarrollo, porque éstos son más subjetivos. Estas opiniones consideramos que pueden ser de carácter general sobre el tipo de prueba de evaluación, pero pueden no tener relación con el hecho de que la prueba de evaluación sea en red o presencial.

El análisis de los blogs de los alumnos nos ha permitido conocer que éstos se encuentran generalmente satisfechos con la experiencia realizada. Uno de los aspectos comunes es que los estudiantes no han encontrado la prueba de evaluación difícil, indicaban además que encontrar la respuesta no era complejo porque los contenidos de la materia se encuentran en red, de ahí que indiquen también que el tiempo ha sido un problema, el hecho de reformular el contenido para responder a la pregunta y de la posibilidad de reflexionar sobre la pregunta hizo que los estudiantes encontraran escasa la hora de realización del examen. El análisis de la información cualitativa y personal de los blogs nos ha permitido complementar y corroborar las conclusiones obtenidas en la encuesta, pues el sentido favorable de las respuestas coincide plenamente.

A partir de los datos mostrados podemos indicar que la aplicación OeLE es un innovador sistema de evaluación en red. La gran ventaja de OeLE es la de que dota al alumno de mecanismos para obtener información sobre sus puntos débiles formativos así como dotar al profesor de mecanismos para conocer el rendimiento individual y colectivo de sus estudiantes para poder reaccionar eficientemente para mejorar el aprendizaje de sus alumnos. Además, es destacable comenzar a valorar la importancia del *feedback*. Nuestro análisis de datos nos ha indicado que el *feedback* (organizado en torno a los fallos de las respuestas de los alumnos) ha sido uno de los aspectos mejor valorados de esta experiencia. Más de la mitad de los alumnos en todos los casos consideran que el *feedback* es algo positivo en entornos de evaluación en red, y además de tener una visión global positiva de este aspecto.

La investigación de entornos de evaluación *on-line* es un aspecto que seguramente será muy tratado en los años venideros. Recordemos también que el Espacio Europeo de Educación Superior, además de la reforma de los títulos universitarios, trata de modificar las metodologías de enseñanza orientando éstas hacia una estrategia basada en el aprendizaje del alumno (De Pablos *et al*, 2007). Sabemos también la relevancia que está adquiriendo el *eLearning* en los proyectos europeos (Rué, 2006). No podemos dejar escapar la oportunidad de estudiar estas nuevas oportunidades desde una perspectiva multidisciplinar y queda una línea abierta de investigación en la cual con muestras más amplias se puedan realizar réplicas de nuestro estudio u otros estudios en esta misma línea que ahonden en las posibilidades de las aplicaciones de web semántica en educación.

5. REFERENCIAS

- Arnal, J., Del Rincón, D. y LaTorre, A. (1992). *Investigación educativa. Fundamentos y Metodología*. Barcelona: Labor.
- Ato, M. y Vallejo, G. (2007). *Diseños experimentales en Psicología*. Madrid: Pirámide.

- Barberá, E. (2006). Aportaciones de la tecnología a la e-Evaluación. En *RED: Revista de Educación a Distancia, monográfico VI*. Murcia: Universidad de Murcia. <http://www.um.es/ead/red/M6/barbera.pdf> [Consultado el 29 de Enero de 2010]
- Berners-Lee, T., Hendler, J. Y Lassila, O. (2001). *The Semantic Web. Scientific American*, 284(5): 34-43, May 2001.
- Cabero, J. (1998). Las aportaciones de las nuevas tecnologías a las instituciones de formación continuas: reflexiones para comenzar el debate, en Departamento de Didáctica y Organización Escolar de la Universidad Complutense-UNED. *Las organizaciones ante los retos del siglo XXI*, pp. 1143-1149. <http://www.desarrollo.upev.ipn.mx/marco/B5TI21.doc> [Consultado el 12 de Septiembre de 2008]
- Cake, M. (2008). Web 1.0, Web 2.0, Web 3.0 y Web 4.0 explained. <http://www.marcuscake.com/key-concepts/internet-evolution> [Consultado el 4 de Mayo de 2010]
- Castells, P. (2005). La Web Semántica. En Bravo, C. y Redondo, M.A. (2005). *Sistemas interactivos y colaborativos en la Web*. Castilla La Mancha: Ediciones de la Universidad de Castilla La Mancha.
- Castellanos, D., Fernández, J., Valencia, R., Martínez, R., Iniesta, M. (2011). Semantic Web Technologies for Supporting Learning Assessment. *Information Sciences*, 181(9), 1517-1537.
- Castillo, S. (2002) (coord.). *Compromisos de la Evaluación Educativa*. Madrid: Prentice Hall.
- Codina, L. (2003). Internet invisible y Web Semántica: ¿el futuro de los sistemas de información en línea?. En *Revista Tradumática. Traducció i Technologies de la Informació y la Comunicació*, n. 2. La Documentació <http://bibliotecnica.upc.es/bustia/arxiu/30094.pdf> [Consultado el 18 de Noviembre de 2010]
- Corbetta, P. (2007). *Metodología y técnicas de investigación social*. Madrid: Mc Graw Hill.
- De Pablos, J. (2004). La formación superior y el reto de las Nuevas Tecnologías de la información, en Martínez, F. y Prendes, M.P. (2004). *Nuevas Tecnologías y educación*. Pearson: Madrid.
- De Pablos, J., Colás, P., González, T. y Jiménez, R. (2007). La adaptación de las universidades al Espacio Europeo de Educación Superior. Un procedimiento metodológico para el diseño de planes estratégicos. *Revista de Investigación Educativa*, 25(2), 533-554.
- Díaz de la Rada, V. (1999). Técnicas de análisis de datos para investigadores sociales: aplicaciones prácticas con SPSS para Windows. Madrid: Ra-Ma.
- Driscoll, M. (2001). Buildind better e-assessment, en *Learning circuits, ASTD*. http://www.astd.org/LC/2001/0601_driscoll.htm [Consultado el 23 de Febrero de 2010]
- Fernández Breis, J.T. y Martínez Béjar, R. (2002) A cooperative framework for integrating ontologies. *Int. J. Hum.-Comput. Stud.*, 56(6): 665-720, 2002.
- Fernández Breis, J.T., Prendes Espinosa, M.P., Castellanos Nieves, D., Martínez Sánchez, F., Valencia García, R. y Ruíz Martínez, J. (2007). Evaluación en e-learning basada en tecnologías de la Web semántica y procesamiento del lenguaje natural. Murcia: Diego Marín.
- Hernández Pina, F. (2001). *Bases Metodológicas de la investigación educativa: I Fundamentos*. Murcia: Diego Marín.
- Alonso, P. (2007). Evaluación formativa y su repercusión en el clima del aula. *Revista de Investigación Educativa*, 25(2), 389-402.

- Perrenhoud, P.H. (1990). *La construcción del éxito y del fracaso escolar*. Madrid: Paidea-Morata.
- Punie, Y., Zinnbauer, D. y Cabrera, M. (2006). A review of the impact of ICT on Learning. Working paper for DG EAC. JRC. Technical notes. <http://ftp.jrc.es/EURdoc/JRC47246.TN.pdf> [Consultado el 21 de Febrero de 2010]
- Prendes, M.P. (2007a). Selección e integración de medios en la enseñanza, en Cabero, J. (2007) (coord). *Tecnología Educativa*. Madrid: Mc Graw Hill.
- Prendes, M.P. (2007b). Internet aplicado a la educación: estrategias didácticas y metodologías, en Cabero, J. (coord.). *Las Nuevas Tecnologías aplicadas a la educación*. Madrid: Mc Graw Hill.
- Rubio, M.J. (2003). Enfoques y modelos de evaluación del e-learning, en *RELIEVE Revista Electrónica de Investigación y Evaluación Educativa*, vol. 9, n. 2, pp. 101-120.
- Esteban Gil, A., Fernández Breis, J.T., Castellanos Nieves, D., Valencia García, R., García Sánchez, F. (2009). Semantic enrichment of SCORM metadata for efficient management of educative contents. *Procedia - Social and Behavioral Sciences, Volume 1, Issue 1*, 927-932.
- Radenkovic, S., Krzavac, N. y Devedzic, V. (2010). An assessment system on the Semantic Web, en Devedzic, V. y Gasevic, D. (ed.), *Web 2.0 and Semantic Web, annals of information systems*. USA: Springer.
- Ramos, G, Suárez, J.M. y Jornet, J. (2006). Validación de un plan de evaluación de teleformación en la empresa. *Revista de Investigación Educativa*, 24(1), 223-238.
- Strother, J.B. (2002). An assessment of the effectiveness of e-learning in corporate training programs, in *The international review of Research in Open and Distance Learning*, vol. 3, n.1. <http://www.irrodl.org/index.php/irrodl/article/viewArticle/83/160> [Consultado el 23 de Febrero de 2010]
- Tejedor, F.J. (1984). *Análisis de varianza aplicado a la investigación en pedagogía y psicología*. Ciencias de la Educación. Madrid: Anaya.

Los datos que se presentan son en parte el resultado de una investigación realizada en el marco del Proyecto 08756/PI/08 "Plataforma Semántica de Formación a la Carta" financiado por la Fundación Séneca.

Fecha de recepción: 16 de marzo de 2011.

Fecha de revisión: 05 de julio de 2011.

Fecha de aceptación: 11 de noviembre de 2011.