

Calibración de ítems vía expertos utilizando Moodle

Experts-based item calibration using Moodle

Concepción Presedo, Ana J. Arméndariz, Javier López-Cuadrado y Tomás A. Pérez
Escuela Universitaria de Ingeniería Técnica Industrial de Bilbao, Universidad del País Vasco (UPV/EHU) España.

Resumen

Calibrar un conjunto de ítems mediante el juicio de expertos es un proceso que consiste en establecer en una métrica común la dificultad de cada ítem partiendo de una muestra de datos que ha sido recogida previamente a individuos especialistas en el tema sobre el que versan los ítems y que contiene estimaciones sobre el nivel de dificultad de cada uno de ellos. Este proceso de calibración no es una tarea simple, pues normalmente implica repartir el conjunto de ítems en distintos tipos de cuestionarios a distribuir entre los expertos, puede ser necesario que algunos ítems se repitan en varios cuestionarios y es conveniente comprobar que el valor estimado de la dificultad se hace sobre ítems no anómalos y se basa en las aportaciones de expertos fiables. Además, se deben consensuar los criterios, a veces dispares, de los expertos consultados.

Este artículo se centra en la construcción y calibración del banco de ítems a partir del juicio de expertos de una manera electrónica y describe un procedimiento para este tipo de calibración utilizando una plataforma educativa web: Moodle. En la actualidad, Moodle es uno de los paquetes de software libre más populares para la creación de cursos y sitios Web basados en Internet, siendo utilizado por instituciones educativas a nivel mundial, para ofrecer formación online.

Palabras clave: expertos | calibración | banco de ítems | Moodle.

Abstract

The calibration of a set of items by expert judgment is the process of establishing the difficulty of each item in a common metric by using a sample of data that has been previously collected from specialists on the subject the items deal with. These data contains estimates of the level of difficulty of each item. This calibration process is not a simple task, since it usually involves the partition of the whole set of items into several questionnaires that are distributed among the experts, as well as, in some cases, the inclusion of common items into several questionnaires. Also it is recommendable to determine that both the items and the obtained estimated values for their difficulty are correct (i.e. non anomalous) and to check that the inputs given by the experts are trustworthy. Furthermore, a consensus must be reached when the criteria of the experts consulted are unequal.

This paper focuses on the item bank construction and its calibration using expert judgments on-line, and describes a procedure for this type of calibration using an e-learning platform: Moodle. Nowadays Moodle is one of the most popular free software packages for creating Internet-based courses and Web sites, and it is used worldwide by educational institutions to provide online training.

Keywords: experts | calibration | item banking | Moodle.

1. INTRODUCCIÓN

Todo sistema de aprendizaje debe complementarse con algún mecanismo de evaluación de los alumnos, que mida la adquisición del conocimiento que se ha producido tras un cierto grado de interacción con el sistema. En un entorno de aprendizaje informatizado, el modo habitual de llevar a cabo esta evaluación consiste en la administración de test a los alumnos ya sea utilizando el formato tradicional o bien empleando test adaptativos. Para garantizar su buen funcionamiento es necesario conocer algunos parámetros que caracterizan las preguntas – como mínimo, la dificultad. El proceso de estimación de estos parámetros se denomina *calibración del banco de ítems* y posee muchas utilidades a la hora de construir test y organizar el dominio de conocimiento de sistemas de aprendizaje.

Los fundamentos para este proceso de calibración se encuentran en las teorías psicométricas, que proporcionan métodos para la construcción de test y proveen de modelos matemáticos que facilitan el análisis de los resultados obtenidos. Históricamente, la primera teoría fue la que a día de hoy se conoce como *Teoría Clásica de los Test* (TCT), y los ítems se calibraban estimando únicamente su dificultad siguiendo la experiencia de individuos doctos en la materia sobre la que trataban. A partir de la aparición de la *Teoría de Respuesta al Ítem* (TRI), surge la calibración estadística (o psicométrica) que consiste en administrar los ítems a una muestra muy amplia de individuos para que los respondan y, a partir de sus respuestas, estimar estadísticamente los valores de los parámetros de la TRI que los caracterizan – dificultad, discriminación y pseudoacierto – y que se tendrán en cuenta o no dependiendo del modelo que se utilice. Una calibración de ítems según la TRI se ejecuta por lo general en varios pasos consecutivos, que están perfectamente detallados en la literatura (López-Cuadrado, 2010; López-Cuadrado, Pérez y Armendariz, 2005; Renom y Doval, 1999). El inconveniente de este nuevo método es que, además de requerir de conocimientos en otras disciplinas (como informática, psicometría y estadística), se precisa de muestras muy numerosas y un procedimiento complejo computacionalmente y con fuertes restricciones a cumplir, lo que constituye un problema a la hora de llevarlo a la práctica (Muñiz, 2010). Por este motivo, aún hoy en día los ítems se siguen calibrando mayoritariamente a partir de estimaciones de expertos.

Calibrar un conjunto de ítems mediante el juicio de expertos es un proceso que consiste en establecer en una métrica común la dificultad de cada ítem partiendo de una muestra de datos que ha sido recogida previamente a indi-

viduos especialistas en el tema sobre el que versan los ítems y que contiene estimaciones sobre el nivel de dificultad de cada uno de ellos. Pudiendo existir múltiples métodos para la recolección de esta muestra de datos, en el contexto de este artículo los datos se recopilarán a través de la distribución de cuestionarios electrónicos a los expertos. El proceso de calibración así definido no es una tarea simple, pues normalmente implica repartir el conjunto de ítems en distintos tipos de cuestionarios a distribuir entre los expertos según el perfil que éstos presenten, puede ser necesario que algunos ítems se repitan en varios cuestionarios y es conveniente comprobar que el valor estimado de la dificultad se hace sobre ítems no anómalos y se basa en las aportaciones de expertos fiables. Además, se deben consensuar los criterios, a veces dispares, de los expertos consultados.

A la hora de informatizar este proceso se ha de tener en cuenta que un cuestionario a administrar entre los expertos para realizar una calibración de ítems no es más que un test con una serie de particularidades, así que toda herramienta capaz de crear y distribuir test es susceptible de ser utilizada para generar cuestionarios. En este sentido, se han desarrollado multitud de paquetes para realizar evaluaciones – como los citados en (Presedo, Armendariz y López-Cuadrado, 2012) – desde programas modestos que generan test convencionales hasta sistemas más complejos y ambiciosos que cubren todo el proceso de desarrollo de un test: construir bancos de ítems, calibrarlos desde diversos modelos de la TRI, efectuar análisis de fiabilidad, y administrar test informatizados tanto clásicos como adaptativos. Un ejemplo es FastTEST (www.assess.com) que, a pesar de su envergadura, es una solución excesivamente rígida, resultando difícil de usar y de hacer compatible con otros módulos ya implementados (Prieto y Delgado, 1999). Esto mismo les ocurre a muchos otros productos software, es decir, algunas de sus características limitan la construcción de test en entornos ya desarrollados con ciertas particularidades propias.

Con objeto de paliar parte de los inconvenientes mencionados, este artículo propone la utilización de plataformas educativas web, que son altamente personalizables y es probable que los individuos interesados en calibrar sus bancos de ítems estén familiarizados con su uso, ya que generalmente serán profesores o personas relacionadas con la enseñanza. Así, la siguiente sección del artículo describe las tareas involucradas en el proceso de calibración de ítems vía expertos y sus posibilidades de informatización. En las secciones tercera y cuarta, se detallan respectivamente la plataforma Moodle y su utilización en el proceso informatizado. Se finaliza con una quinta sección de conclusiones.

2. EL PROCESO GLOBAL DE CALIBRACIÓN DE ÍTEMS Y SU INFORMATIZACIÓN

Una vez conocido el banco de ítems que se desea calibrar, todo proceso global de calibración consta de dos fases consecutivas: *administración de ítems* y *análisis de datos y calibración* que a su vez se pueden descomponer en otra serie de tareas también consecutivas (López-Cuadrado, 2010). A continuación, se describen estas tareas para calibraciones vía expertos, adaptadas de Arruabarrena (2010).

En la primera fase del proceso – *Administración de ítems* – se elige y prepara un sistema para la administración de cuestionarios, seguidamente se distribuyen entre los expertos para su cumplimentación y se almacenan sus respuestas – normalmente en una base de datos – para su posterior análisis. Para poder ejecutar esta fase, una vez especificado el banco de ítems, se debe dejar constancia de varias decisiones de diseño sobre: el objetivo de la fase (en cuanto a duración máxima o volumen de datos a recoger por ítem), el conjunto de expertos de que dispondrá (en cuanto a su captación, identificación y establecimiento de su perfil) y los cuestionarios a distribuir (en cuanto a la estructura e ítems de cada uno y a los datos que se pedirán al experto por cada ítem a calibrar).

120

En la segunda fase – *Análisis de datos y calibración* – se lleva a cabo el análisis en cuanto a la fiabilidad de los datos recogidos y la obtención de los valores del parámetro dificultad de cada ítem a partir de las aportaciones de los expertos recogidas mediante la administración. Las decisiones previas de diseño que posibilitan la ejecución de esta fase son dos: por un lado, determinar todos los procedimientos que tratan de atestiguar la calidad de los ítems y el rigor con el que han participado los expertos consultados una vez recogidos los datos, para incluirlos o no en el cálculo de la dificultad; y por otro lado, indicar los detalles de cálculo para hallar la dificultad final de cada ítem. En este sentido, las teorías psicométricas proporcionan distintos métodos para analizar la muestra recogida y estimar la dificultad de cada ítem. Estos métodos de análisis permiten verificar la calidad tanto de los ítems como de los test diseñados, y son especialmente útiles en los casos en los que no sea posible revisar los ítems antes de administrar los test. En concreto, la TCT se fundamenta en el denominado modelo clásico, que establece una relación lineal entre la habilidad del examinado y la puntuación obtenida en el test realizado. En un *análisis clásico*, se puntúa a los individuos con el número de respuestas correctas – puntuando cada respuesta como 1 si es correcta ó 0 si es incorrecta – y se hallan una serie de estadísticas

importantes, que basan sus cálculos en proporciones, correlaciones y estadísticas descriptivas tales como la media, desviación típica, varianza, mínimo y máximo, y que incluyen estadísticas a nivel de ítem y estadísticas a nivel de test (Muñiz, 2000). El análisis clásico se utiliza en contraposición al enfoque dado por la TRI que es mucho más potente pero solamente funciona con un tamaño de muestras del orden de cientos o mayores, es decir, la TRI solamente es adecuada para analizar una muestra cuando se dispone de evaluaciones a gran escala (Muñiz, 2010). Así, los procedimientos y cálculos a describir se pueden inspirar en los análisis clásicos de la TCT, o bien ser estadísticos o métodos particulares para cada caso (Arruabarrena, 2010; Armendariz, 2014). Cabe destacar que aunque los métodos ideados para analizar las respuestas recogidas sean simples, el análisis manual de los datos no es fácil y en muchas ocasiones será preciso el uso del ordenador.

La última tarea del proceso de calibración es *construir el banco de ítems calibrado*. Un banco de ítems calibrado (Umar, 1997) es una colección de ítems, que está almacenada de manera estructurada junto con sus características, tanto de contenido como psicométricas, y cuyos parámetros han sido estimados y transformados a una escala común mediante un proceso de equiparación. La información que se guarda en el banco suele incluir, por cada ítem, un identificador único, el enunciado y sus materiales asociados (tablas, fotos, videos, sonidos, etcétera), las opciones correctas e incorrectas, una referencia al dominio específico sobre el que evalúa, el número de veces que ha sido administrado y los parámetros del modelo de la TRI al que corresponden. También se suele disponer de índices y tablas que agilicen la búsqueda del ítem más apropiado a cada necesidad e incluso datos de carácter más general como el nombre del autor o la fecha de creación de cada ítem.

En cuanto al software que se puede utilizar para informatizar este proceso global de calibración, a día de hoy existen multitud de paquetes informáticos, tanto para la construcción y administración de ítems y cuestionarios como para tareas concretas del análisis psicométrico y de la calibración de un banco de ítems. En primer lugar, cualquiera de las herramientas mencionadas en la introducción – incluidas las plataformas educativas web – puede ser utilizada como *sistema para la construcción y administración de cuestionarios*. En segundo lugar, puede encontrarse *software de apoyo para el análisis estadístico de la muestra de datos recogidos*. Como software para realizar análisis personalizados de este tipo se puede emplear cualquier paquete estadístico del mercado, por ejemplo SPSS (www.ibm.com/software/analytics/spss) o R (www.r-project).

org); y para añadir más potencia gráfica puede ser útil cualquier programa que manipule datos mediante hojas de cálculo, como Microsoft Excel u OpenOffice (www.openoffice.org). Esta última idea ha generado toda una serie de *sistemas autocontenidos*, es decir, aplicaciones en las que el usuario registra en una o varias hojas de cálculo los datos con los resultados del test (las respuestas de los individuos junto con las respuestas correctas) y las especificaciones de los análisis; cuando selecciona las distintas opciones disponibles y se ejecutan, las estadísticas se actualizan en tiempo real y se crean los informes correspondientes en forma de tablas y gráficos. El propósito de estos informes es ayudar a los que utilizan test a evaluar la calidad de sus ítems y de los propios test, mediante el examen de sus características psicométricas. Un ejemplo de este tipo de software es CITAS (Classical Item and Test Analysis Spreadsheet), que realiza automáticamente el análisis clásico de la TCT, y que está disponible gratuitamente en www.assess.com. En cuanto al *software de apoyo para la estimación del parámetro dificultad*, si el método ideado se basa en estadísticas descriptivas sobre la muestra recogida se puede utilizar cualquiera de los programas ya mencionados. Otro enfoque totalmente diferente al anterior, es el que intenta estimar la dificultad de cada ítem buscando un consenso online entre los distintos expertos que participan en la calibración del parámetro, para lo cual puede ser útil el empleo de *herramientas colaborativas* similares a las que ofrecen la mayoría de las plataformas educativas Web.

3. LAS PLATAFORMAS EDUCATIVAS WEB Y MOODLE

Una *plataforma educativa web* es una aplicación que integra un conjunto de herramientas para la enseñanza-aprendizaje en línea, permitiendo una enseñanza no presencial (*e-learning*) y/o una enseñanza mixta (*b-learning*), donde se combina la enseñanza por Internet con experiencias en la clase presencial. Estas plataformas ofrecen la posibilidad de proporcionar conocimiento en cualquier momento y en cualquier lugar donde haya una conexión a Internet, y su uso ha transformado una gran parte de los espacios de enseñanza tradicionales en espacios virtuales de enseñanza y aprendizaje o EVA. Actualmente el software más utilizado para crear y dar soporte a múltiples EVA son los *sistemas de gestión del aprendizaje* o LMS (Learning Management System). La creación de estos EVA, normalmente, se realiza utilizando una plantilla que personaliza un conjunto de herramientas que el diseñador considera necesarias para llevar a cabo el proceso de aprendizaje (López Alonso y Matesanz del Barrio, 2009) como: (1) herramientas de administración del EVA, (2) herramientas de comuni-

cación, (3) herramientas de gestión de grupos, (4) herramientas de evaluación y (5) herramientas de gestión de contenidos.

Las *herramientas de administración del EVA* permiten realizar las operaciones de creación, borrado y modificación de usuarios del sistema, definir roles, asignar tutores, y llevar un control de los accesos al sistema. También permiten la creación, borrado y modificación de los EVA que se soportan. Las *herramientas colaborativas o de comunicación* permiten la interacción entre los diferentes roles que hay en un entorno de aprendizaje, como profesores, tutores o estudiantes. Las herramientas de comunicación se dividen según el tipo de interacción temporal en herramientas de comunicación síncrona – si se establece una comunicación en tiempo real entre los participantes – o asíncrona. Algunos ejemplos de herramientas síncronas son chat, pizarra electrónica, audio y videoconferencias; y como herramientas asíncronas se pueden citar el correo electrónico, los foros y las wikis. Las *herramientas de gestión de grupos* permiten realizar las operaciones de alta, modificación o borrado de grupos de alumnos y la creación de escenarios virtuales para el trabajo cooperativo de los miembros de un grupo. Estos escenarios de grupo incluyen carpetas para el intercambio de archivos, herramientas para la publicación de los contenidos, y foros o chats privados para los miembros de cada grupo. Las *herramientas de evaluación* permiten la creación, edición y realización de pruebas de evaluación. Las pruebas pueden ser de autoevaluación, de tal forma que el estudiante pueda afianzar o corregir posibles problemas que se le han presentado en el aprendizaje. También pueden permitir al profesor obtener un informe de las respuestas que han dado los estudiantes. Las *herramientas de gestión de contenidos* también conocidas como *LCMS* (Learning Content Management System) tienen la labor de crear los cursos y sus contenidos en forma de módulos que se pueden personalizar, manejar y usar en diferentes ocasiones. Para ello, proporcionan una gestión de contenidos orientada al e-learning integrando generalmente estándares de producción de contenidos educativos reutilizables (como los proporcionados por IMS, AICC y/o SCORM).

Una de las plataformas más completas y populares en el ámbito educativo es Moodle (moodle.org) que dispone de gran cantidad de documentación tanto en Internet (docs.moodle.org) como en la literatura (Büchner, 2011; Sánchez Rojo, 2010). *Moodle* es un software de código abierto y modular que además

de todas las funcionalidades de un LCMS ofrece varios métodos de autenticación y servicio propio de mensajería entre los usuarios. Los datos que maneja Moodle se almacenan en el servidor de la plataforma, en una base de datos denominada *moodle*, y se organizan en diferentes estructuras. Las principales estructuras que utiliza Moodle son: *usuarios*, *cursos* y *módulos*. A su vez, los módulos se pueden clasificar en tres grupos: *recursos*, *actividades* y *bloques*.

En cuanto a los *usuarios* de la plataforma, Moodle mediante su servicio de autenticación, proporciona una base de datos interna que respalda la infraestructura de acreditación de usuarios en la plataforma, es decir, la creación de sus cuentas de acceso a Moodle. Estos usuarios se diferencian atendiendo a un *sistema de roles* que determinan sus posibilidades de interactuar con la plataforma. Cada rol asigna una serie de *privilegios* que permiten restringir ciertas características o funcionalidades a un usuario según su nivel. Existen las siguientes categorías de usuarios, ordenadas de mayor a menor, en función del nivel de privilegios: administrador principal, gestor, creador de curso, profesor, profesor sin derecho de edición, estudiante e invitado. Concretamente, dentro de un curso el rol de *profesor (Teacher)* tiene acceso completo para construirlo, editarlo y matricular a otros usuarios, puede añadir tanto recursos como actividades, y cambiar su configuración; y el rol de *estudiante (Student)* permite participar en las actividades y descargar materiales del curso. Aparte de uno o varios roles, cada usuario tiene asociado un *perfil* formado por información personal en la que obligatoriamente deben aparecer sus datos de acceso a Moodle, un nombre, apellido(s), dirección de correo y ciudad.

124

Respecto a los *cursos* son los elementos principales de Moodle, y constituyen su unidad básica de organización. Inicialmente, todos los cursos que se crean están clasificados por categorías. Dentro de ellos se pueden encontrar los recursos y/o las actividades. Los recursos están orientados a procesos de aprendizaje pasivos para cuestiones de auto-estudio, y priorizan la interacción persona-contenido (trabajos individuales con el contenido). Un *recurso* es un documento, imagen, página web, etc. que sirve como soporte a las actividades del curso, por ejemplo, instrucciones o información sobre alguna sección del curso. Moodle considera diferentes tipos de recursos: una página de texto, una página Web, un enlace a un archivo, un enlace a una Web o una etiqueta. Por su parte, una *actividad* se orienta a procesos de aprendizaje activos, y prioriza la interacción entre personas (de una a una y de varias a varias). Moodle ofrece gran cantidad de actividades como chats, foros, glosarios, blogs, wikis, sistemas de control de actividades de aprendizaje, talleres y cuestionarios (o quizzes). De este modo, se permite

crear exámenes o cuestionarios. Un *cuestionario* no es más que un conjunto de preguntas, con la salvedad de que se incluye su puntuación. Las preguntas así definidas se almacenan en un *banco de preguntas* a nivel de curso y cada una de ellas es de alguno de los siguientes tipos: numérica, calculada, opción múltiple, opción múltiple calculada, respuesta corta, ensayo, emparejamiento, emparejamiento de respuesta corta aleatoria, verdadero/falso, descripción y respuestas incrustadas. A nivel de cuestionario, permite configurar la puntuación final así como el modo de navegar por las preguntas y es capaz de dar realimentación (inmediata o diferida, por ítem o una vez intentado todo el test) al alumno que lo intenta. Gracias a las puntuaciones definidas Moodle es capaz de calificar automáticamente los cuestionarios intentados y permite visualizar tanto las respuestas dadas por cada alumno que intentó el cuestionario, como las *puntuaciones* obtenidas en cada apartado y los tiempos consumidos. Todos estos datos se pueden descargar directamente en *informes* de resultados con varios formatos – hoja de cálculo Excel, hoja de cálculo OpenOffice, archivo en texto plano y archivo en XML.

Por último, antes de comenzar la impartición de un curso es necesario: crear y configurar el curso, crear y configurar las actividades del curso que el alumno deberá realizar y matricular a los participantes en el curso con un determinado rol. Las distintas opciones de configuración se encuentran en la última de las estructuras principales: los bloques. Los *bloques* de Moodle proporcionan información o funcionalidad adicional al estudiante o al profesor como un calendario, el bloque de administración de un curso, usuarios en línea, novedades, canales RSS remotos, etc. Estos bloques aparecen en los laterales de sus páginas y pueden agregarse, borrarse, desplazarse, etc. El más importante es el *bloque de Administración*, que permite configurar el LMS y gestionar todos los servicios y herramientas disponibles como cursos, actividades, puntuaciones, informes, recursos, servicio de autenticación, usuarios, formato de la página principal, apariencia e idioma del sitio Web, servicios Web y mensajería. En lo referente a la *configuración del LMS*, el administrador principal puede tomar una serie de decisiones y personalizar la plataforma en función del uso que se le vaya a dar. Esto significa que activará o no ciertas opciones disponibles en la plataforma y establecerá valores por defecto a nivel global, que después podrán ser modificadas por otros usuarios en sus cursos siempre que su nivel de privilegios se lo permita. Como ejemplos, puede decidir cuál será el método de autenticación o el idioma por defecto, personalizar los datos que se piden en el perfil de las cuentas de usuario, modificar los privilegios de algún rol, y

activar el servicio de mensajería si decide permitir el envío de mensajes entre distintos usuarios.

4. EL PROCESO DE CALIBRACIÓN DE ÍTEMS UTILIZANDO MOODLE

Cuando se utiliza una plataforma educativa Web como Moodle, el proceso global de calibración implica crear y configurar un curso que contenga los cuestionarios diseñados y matricular en él a los distintos participantes que serán de dos tipos: los *expertos* y el *responsable* de la calibración, entendido éste último como el individuo que desea obtener los ítems calibrados y que realizará todas las tareas necesarias para tal fin. Una vez preparado el curso para su impartición se suministrarán los cuestionarios a esos expertos y se recogerán los datos con los que se realizará la fase de análisis y calibración de los ítems. Todas las tareas que el responsable debe llevar a cabo se han concretado en una serie de 10 pasos que se detallan a continuación.

126

Paso 1 – Dar de alta en Moodle tanto al responsable como a los expertos participantes. Consiste en solicitar al administrador cuentas de acceso a la plataforma para los distintos participantes. Cuando se crea una cuenta de usuario de Moodle se hace con un determinado rol, generalmente el de profesor (*Teacher*) o el de estudiante (*Student*).

Paso 2 – Crear un curso vacío y matricular al responsable. Para que el responsable de la calibración pueda realizar este paso deberá tener el rol de creador de curso, en caso contrario deberá solicitarlo también al administrador. En cuanto a su rol dentro del curso, la labor del responsable es doble: por un lado debe tomar todas las decisiones de diseño y por otro debe supervisar, coordinar y controlar el desarrollo del proceso una vez comenzada su ejecución. Así, el responsable de la calibración debe tener capacidad para crear los cuestionarios y gestionar todo lo relacionado con ellos. Por ejemplo, debe ser capaz de incluir en el curso las instrucciones de cumplimentación, por lo que debe tener permiso para añadir material adicional al curso, rellenar sus datos personales, y monitorizar en todo momento el progreso de la administración. Además resulta conveniente que el responsable pueda establecer plazos tanto para la disponibilidad del curso como para la cumplimentación y envío de cuestionarios por parte de los expertos. También puede ser deseable que pueda matricular a nuevos expertos en el curso, si quiere recoger más respuestas. Por último, el responsable debe tener la opción de mandar mensajes a los expertos cuando

tenga todo listo o cuando se vaya a terminar el plazo previsto, es decir, se debe permitir al responsable comunicarse con los participantes mediante avisos. Por ello, en Moodle bastará con que el responsable esté matriculado en el curso con el rol de profesor.

Paso 3 – Configurar el curso. Consiste en elegir las opciones más adecuadas para el curso, entre ellas el formato, el idioma y las fechas en que estará disponible.

Paso 4 – Crear en el curso los cuestionarios y el resto del material a distribuir. Consiste en crear el banco de ítems y los cuestionarios. A la hora de diseñar los distintos cuestionarios, por cada ítem a calibrar se han de pedir al experto como mínimo dos datos: obligatoriamente el nivel de dificultad estimado y opcionalmente la respuesta al ítem (útil como información de control). Así, por cada ítem a calibrar se crearán dos preguntas asociadas en Moodle: una con el propio ítem del banco y otra en la que se solicite el nivel y cuya respuesta correcta será cualquiera que pertenezca al rango de dificultades disponibles. Para cada pregunta habrá que indicar su tipo, respuesta correcta y puntuación (en formato 1/0). A continuación se crearán tantas actividades como tipos de cuestionarios se hayan definido. Una vez creado cada cuestionario, el responsable puede previsualizarlo y probar su funcionamiento. También es conveniente incluir en el curso instrucciones de cumplimentación en las que se indique claramente cuál es la escala de valoración a la hora de estimar la dificultad con objeto de que todos los expertos apliquen el mismo criterio, lo que evitará la necesidad de una posterior equiparación de puntuaciones (Arruabarrena, 2010). Para ello se puede crear uno o varios documentos y añadirlos al curso en forma de recursos Moodle.

La realización de esta labor de modo manual es bastante tediosa por lo que puede ser conveniente estudiar las posibilidades de importación de Moodle desde otro tipo de programas. En este sentido, otra opción para crear el curso junto con sus materiales es seguir algún estándar de producción de contenidos para *crear un curso reutilizable desde otra herramienta e importarlo* a Moodle. En cuanto a la cobertura de estándares, Moodle permite importar paquetes SCORM y soporta diversos estándares de IMS, posibilitando, por ejemplo, la importación de cursos en el formato estándar IMS CC. IMS utiliza su especificación IMS QTI (Smythe, Shepherd, Brewer y Lay, 2002) para la representación de ítems y test. Aunque los entresijos son complicados, existen herramientas que permiten la creación, modificación y gestión de cuestiones y test en IMS QTI. Una de ellas es Respondus (<http://www.respondus.com/>). Con los ítems y test así creados

se puede generar un curso reutilizable, mediante la creación de un paquete de contenido IMS, que se podrá llevar a cualquier otra aplicación software que soporte su importación. La estructura de un paquete de contenidos IMS se compone de dos elementos principales: los contenidos locales a empaquetar (los ítems, cuestionarios y/u otros archivos relevantes para la administración, por ejemplo las instrucciones) y un manifiesto de nivel superior que los describe y organiza dentro del curso. IMS trabaja con dos estándares: el paquete de contenidos IMS CP (Smythe y Nielsen, 2007) y el cartucho común o IMS CC (Riley y Mills, 2008), que representan dos métodos de distribución de cursos reutilizables. Para facilitar la creación y modificación de este tipo de paquetes, han aparecido un conjunto de editores que permiten realizar estas operaciones de manera visual e intuitiva. Uno de los más utilizados es RELOAD (<http://www.reload.ac.uk/>). Adicionalmente IMS ofrece una herramienta de validación en la web de cartuchos de contenido IMS CC (<http://validator.imsglobal.org/cc/>) que indica si el cartucho ha sido creado correctamente.

128

Paso 5 – Configurar los cuestionarios. Consiste en elegir las opciones más adecuadas para una administración de este tipo. Concretamente, lo normal es que el responsable permita que un experto pueda realizar cada cuestionario que se le asigne en varios intervalos, es decir, se consientan varios intentos en los que se conserve lo ya contestado anteriormente y con un modo de navegación no lineal para que pueda rellenar las preguntas en el orden que prefiera y sin límites de tiempo. También es importante no ofrecer realimentación sobre la respuesta correcta hasta que no haya terminado el cuestionario. En cuanto a su formato de presentación en Moodle es conveniente que aparezcan en una misma página todos los datos que se piden por cada ítem a calibrar.

Paso 6 – Matricular en el curso a los expertos. Dentro del curso la labor de los expertos consiste en estimar el nivel de dificultad de cada ítem mediante la cumplimentación de los cuestionarios. Así, el experto administrado debe tener permisos para acceder al material del curso, realizar y enviar los cuestionarios, comunicarse con el responsable y actualizar sus datos personales. Por tanto, deberá ser matriculado en el curso con el rol de estudiante.

Paso 7 – Dar visibilidad al curso. Para que un curso Moodle esté disponible para sus estudiantes, que en este caso son los expertos matriculados, el responsable de la calibración debe marcar la opción *visible* en las opciones del curso.

Paso 8 – Administrar los cuestionarios. Consiste en impartir el curso creado de la misma forma que se impartiría cualquier otro. Para facilitar esta labor, el responsable puede establecer grupos con los distintos expertos que tienen que rellenar cada tipo de cuestionario. Se puede controlar automáticamente, por ejemplo, si un cuestionario está disponible o no, mediante el rastreo del grado de finalización en el nivel curso y el acceso condicional, que permiten fijar las condiciones – basadas en la fecha, la calificación o el grado de finalización – que lo controlan. En cuanto al *almacenamiento de datos para la fase de análisis de datos y calibración* el responsable no debe hacer nada ya que, durante la impartición del curso, Moodle habrá recogido automáticamente en su base de datos toda la información aportada por los expertos, tanto los datos personales como las respuestas dadas.

Paso 9 – Realizar la fase de análisis de datos y calibración. Consiste en que el responsable compruebe la calidad de los datos recogidos en la tarea anterior y calcule el parámetro dificultad para cada ítem del banco con los datos que superen los criterios establecidos. La plataforma permite realizar análisis manuales eliminando aportaciones de expertos y/o ítems no fiables y también puede establecer actividades – por ejemplo un foro – para consensuar online la dificultad de cada ítem en conflicto. Sin embargo, Moodle no contempla el uso de métodos estadísticos sobre los datos almacenados. Una solución para este caso, es que el responsable elija otra herramienta software de apoyo capaz de soportar alguno de los formatos de exportación que estén habilitados para los informes generados automáticamente por Moodle. Esto le permitirá descargarse los informes que contienen los datos que necesite dependiendo de los análisis y cálculos que vaya a realizar, en un formato adecuado – por ejemplo Excel, que es soportado por la mayoría de programas estadísticos – para su posterior tratamiento con la herramienta seleccionada.

Paso 10 – Construir el banco de ítems calibrado. Consiste en transformar el banco de preguntas en un banco de ítems calibrado reflejando en Moodle los resultados obtenidos en la calibración, es decir, el valor del parámetro dificultad de cada ítem. La opción más fácil que ofrece Moodle es crear en el banco de preguntas tantas categorías como niveles de dificultad hayan surgido e incluir en cada una de ellas los ítems calibrados correspondientes, que ya están en la plataforma. Así, el profesor dispondrá de los ítems clasificados por dificultad a nivel de curso, que se podrán utilizar en exámenes a los alumnos.

5. CONCLUSIONES

En este artículo se ha descrito el proceso global de calibración de ítems y se ha establecido un método práctico en 10 pasos para realizar calibraciones de bancos de ítems vía expertos utilizando la plataforma e-learning Moodle. Este método proporciona múltiples ventajas, entre ellas que el docente puede utilizar una única herramienta informática tanto para impartir sus cursos como para calibrar los ítems que empleará en sus exámenes. Además, puede emplear aplicaciones software gratuitas en todo el proceso, como sería la combinación Moodle, OpenOffice y R. Por su parte, el uso de estándares e-learning – en este caso los de IMS, que es una de las iniciativas más populares con más de 20 estándares disponibles de forma gratuita en Internet – permite la creación de contenidos educativos reutilizables y facilita la interoperabilidad entre distintos sistemas. Por otro lado, el método propuesto no es exclusivo para Moodle, sino que se podría explorar la posibilidad de administrar los ítems a los expertos mediante otras plataformas educativas similares, como por ejemplo Blackboard, Atutor, Claroline, Ilias o Sakai.

130

Los autores han probado con éxito la viabilidad de las ideas expuestas utilizando Moodle y Microsoft Excel para replicar un caso real de calibración (Arruabarrena, 2010; López-Cuadrado, Pérez, Vadillo y Gutiérrez, 2010) que se realizó para un banco de 252 ítems de selección múltiple que evalúan la capacidad gramatical del alumno en el idioma vasco.

REFERENCIAS

- Armendariz, A. J. (2014). *Métodos de desarrollo dirigidos por modelos y workflows para la calibración psicométrica de ítems: El sistema CALLIE*. (Tesis doctoral), Universidad del País Vasco (UPV/EHU), Bilbao.
- Arruabarrena, R. (2010). *E-learning y la calibración de ítems de test: Teoría de Respuesta al Ítem versus calibración basada en juicios de expertos. Un estudio empírico*. (Tesis doctoral), Universidad del País Vasco (UPV/ EHU), San Sebastián.
- Büchner, A. (2011). *Moodle 2 Administration*. Birmingham (UK): Ed. Packt Publishing Ltd.
- López-Cuadrado, J. (2010). *Test adaptativos informatizados de ingreso en un sistema e-learning*. Saarbrücken: LAP Lambert Academic Publishing.
- López-Cuadrado, J., Pérez, T. A. y Armendariz, A. J. (2005). Evaluación mediante test: ¿Por qué no usar el ordenador? *Revista Iberoamericana de Educación*, 36(11), 1-15. <http://www.rieoei.org/deloslectores/1040Lopez.PDF>

- López-Cuadrado, J., Pérez, T. A., Vadillo, J. A. y Gutiérrez, J. (2010). Calibration of an item bank for the assessment of Basque language knowledge. *Computers & education*, 55(3), 1044-1055.
- López Alonso, C. y Matesanz Del Barrio, M. (2009). *Las plataformas de aprendizaje: del mito a la realidad*. Madrid: Biblioteca Nueva.
- Muñiz, J. (2000). *Teoría clásica de los tests*. Madrid: Ediciones Pirámide.
- Muñiz, J. (2010). Las teorías de los tests: Teoría Clásica y Teoría de Respuesta a los Ítems. *Papeles del Psicólogo*, 31(1), 57-66.
- Presedo, C., Armendariz, A. J. y López-Cuadrado, J. (2012). *Calibración de ítems para test informatizados: Descripción detallada de las fases en la construcción de test de evaluación adaptativos mediante ordenador*. Saarbrücken: EAE Editorial Académica Española.
- Prieto, G. y Delgado, A. R. (1999). Medición cognitiva de las aptitudes. En J. Olea, V. Ponsoda y G. Prieto (Eds.), *Tests informatizados: fundamentos y aplicaciones* (pp. 207-226). Madrid (España): Ediciones Pirámide.
- Renom, J. y Doval, E. (1999). Tests adaptativos informatizados: estructura y desarrollo. En J. Olea, V. Ponsoda y G. Prieto (Eds.), *Tests informatizados: fundamentos y aplicaciones* (pp. 127-162). Madrid (España): Ediciones Pirámide.
- Riley, K. y Mills, D. (2008). *IMS Common Cartridge (CC) Authorization Web Service v1.0 Final Specification*. Recuperado desde www.imsglobal.org/cc/
- Sánchez Rojo, I. (2010). *Plataforma educativa Moodle. Administración y gestión*. México: Alfaomega/Ra-Ma.
- Smythe, C. y Nielsen, B. (2007). *IMS Content Packaging (CP) Specification Primer v1.2 Public Draft*. Recuperado desde www.imsglobal.org/content/packaging/
- Smythe, C., Shepherd, E., Brewer, L. y Lay, S. (2002). *IMS Question & Test Interoperability: ASI Information Model Specification. Final Specification Version 1.2*. Recuperado desde www.imsglobal.org/question/
- Umar, J. (1997). Item banking. En J. Keeves (Ed.), *Educational research, methodology, and measurement (2nd edition)* (pp. 923-930). Oxford: Pergamon.

