

# Sistema de ayuda para la calibración de ítems por el procedimiento basado en el juicio de expertos

Concepción Presedo, Universidad del País Vasco / Euskal Herriko Unibertsitatea, España  
Anaje Armendariz, Universidad del País Vasco / Euskal Herriko Unibertsitatea  
Javier López-Cuadrado, Universidad del País Vasco / Euskal Herriko Unibertsitatea  
Tomás A. Pérez, Universidad del País Vasco / Euskal Herriko Unibertsitatea

**Resumen:** La evaluación es un aspecto esencial del aprendizaje, siendo los bancos de ítems un recurso común para realizar este tipo de quehaceres. La evaluación profesional utiliza bancos calibrados, siendo la calibración un proceso para la estimación de parámetros de dichos ítems, por ejemplo, la dificultad. Una opción para la calibración de ítems es el proceso basado en utilizar un panel de expertos para recabar sus valoraciones y determinar a partir de ellas los parámetros. Este artículo describe CALLIE-PRO: un sistema que personaliza y automatiza el proceso de calibración de ítems basado en el juicio de expertos. El sistema guía al usuario recomendando una opción para cada decisión que hay que tomar, adaptándose a alteraciones en la especificación y avisando de decisiones no recomendables o incorrectas. Además, a partir de la especificación del proceso, genera un paquete de contenidos de e-learning que se puede incluir en un sistema de gestión del aprendizaje (LMS) como, por ejemplo, Moodle. El paquete se encarga de recopilar los juicios de expertos a través de la red y finalmente, genera un workflow ejecutable con las actividades que se deben realizar utilizando los datos recopilados, generalmente estudios y cálculos, para estimar la dificultad de los ítems.

**Palabras clave:** ítems, calibración, expertos, moodle

**Abstract:** Evaluation is a key issue in learning and item banks are very common resources used to perform activities related to it. Professional evaluation deals with banks that are calibrated, which means that the parameters of the items (for example, their difficulty) have previously been estimated. One option for item calibration consists in gathering the assessments of a panel of experts and using them to determine the values of the item parameters. This paper describes CALLIE-PRO: a system that customizes and automates the process of item calibration based on expert judgment. It guides the user by recommending a choice for every decision to be made, adapting to changes in the specification, and warning of non advisable or incorrect choices. It also creates an e-learning content package that depends on the process specification and can be included in a learning management system such as, for example, Moodle. This package is responsible for gathering the assessments of the experts via the internet and generating an executable workflow that will include the activities to be performed, which are mainly analyses and calculi, by using all gathered data. Finally, the system provides estimates for the parameters (difficulty) of the items.

**Keywords:** Items, Calibration, Experts, Moodle

## 1. Introducción

Todo sistema de aprendizaje debe complementarse con algún mecanismo de evaluación de los alumnos, que por lo general sirve para medir la adquisición del conocimiento que se ha producido tras un cierto grado de interacción con el sistema. 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.

Los test adaptativos informatizados (TAI) son test que emulan el comportamiento de un evaluador humano, y que se generan dinámicamente basándose en las respuestas que va dando el examinado. La idea es que si el evaluado falla una pregunta, la siguiente cuestión que se le plantee será más fácil; y viceversa, como consecuencia de un acierto se administrará un ítem ligeramente más difícil. Los TAI ofrecen multitud de ventajas frente a los test tradicionales sobradamente discutidas en la literatura (Kingsbury y Weiss 1983; Wainer, Dorans, Flaugher, Green, Mislevy, Steinberg y



Thissen 2000), aunque para garantizar su buen funcionamiento es necesario estimar los parámetros del modelo de la Teoría de Respuesta al Ítem (TRI) – como mínimo, la dificultad – que caracterizan cada pregunta. Este proceso de estimación se denomina calibración del banco de ítems.

Históricamente los ítems se han calibrado estimando únicamente su dificultad siguiendo la experiencia de individuos doctos en la materia sobre la que tratan. Con la aparición de la 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 (generalmente dificultad, discriminación y pseudoacierto). Si se usa este último método, además de requerir conocimientos en otras disciplinas (como psicometría y estadística) que los creadores de nuevo material no tienen por qué dominar, se precisa de muestras muy numerosas y un procedimiento computacionalmente complicado, lo que constituye un problema a la hora de llevarlo a cabo. Por este motivo, aún hoy en día los ítems se siguen calibrando mayoritariamente a partir de estimaciones de expertos (Muñiz 2010).

El trabajo expuesto se engloba dentro de la investigación desarrollada por el Grupo de Hipermidia y Multimedia (GHyM) ligado al Departamento de Lenguajes y Sistemas Informáticos de la Universidad del País Vasco. Este grupo trabaja en el área del e-learning desde 1994, cuando empezó a diseñar Hezinet, un sistema hipermidia adaptativo (SHA) para el aprendizaje de la lengua vasca, basado en la realización de actividades implementadas mediante ítems (Pérez 2000) que ha sido instalado en más de 150 centros del País Vasco. Desde entonces, el grupo se ha centrado en el estudio de este SHA con el objetivo de realizar mejoras continuas sobre la aplicación y, en algunos casos, su evolución a una nueva versión. Así, en 2002 se abrió una línea de investigación centrada en el mecanismo de evaluación para el ingreso de los alumnos en el sistema, esto es, la evaluación que se efectúa a cada nuevo estudiante con objeto de determinar su nivel de inicio en Hezinet.

Durante el transcurso de estas investigaciones se renovó la arquitectura del sistema Hezinet dando lugar a una nueva arquitectura: el hiperentorno adaptativo de aprendizaje genérico ELSA (Armendariz, López-Cuadrado, Tapias, Villamañe, Sanz-Lumbier y Sanz-Santamaría 2003). ELSA proporciona una infraestructura e-learning, basada en servicios web sobre la plataforma .Net de Microsoft, que cumple con los estándares internacionales IMS QTI (Smythe, Shepherd, Brewer y Lay 2002) y ADL SCORM para representar el dominio didáctico. Se trata de una arquitectura orientada a la evaluación, que incorpora como nuevos elementos el módulo de evaluación y el módulo de calibración.

Por otro lado, ambas vías de investigación requirieron la revisión previa y calibración de un banco de ítems de evaluación para el test de ingreso que han servido de base para crear el sistema de clasificación de alumnos de Boga ([www.ikasten.ikasbil.net](http://www.ikasten.ikasbil.net)), la versión en Internet de Hezinet. Primero, un grupo de expertos estimó el parámetro dificultad de los ítems cumplimentando distintos cuestionarios en papel y después, tomando como base las estimaciones anteriores para diseñar los subtest a administrar, se realizó la calibración estadística que propugna la TRI. Así, se consiguieron dos calibraciones del banco de ítems mediante dos métodos diferentes: mientras que la primera estableció el valor de la dificultad para cada ítem a partir de la información proporcionada por los expertos, la segunda de ellas les asignó los valores de los tres parámetros de la TRI. Como resultado de la experiencia se dotó a Hezinet de un módulo de evaluación que genera TAIs de ingreso al sistema (López-Cuadrado 2010) y se formalizó un único procedimiento integral para la calibración de ítems (Arruabarrena 2010) que puede instanciarse para ser utilizado en cualquiera de los dos tipos de calibración. Este artículo, no obstante, se centrará en las calibraciones de ítems siguiendo el criterio de expertos.

En el contexto de este artículo 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 normalmente precisa repartir el conjunto de ítems en distintos cuestionarios a administrar, 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.

El procedimiento integral de calibración considera que el banco de ítems a calibrar estará formado exclusivamente por ítems de selección múltiple en formato texto para prevenir problemas de multidimensionalidad (López-Cuadrado, Pérez, Vadillo y Gutiérrez 2010). Una vez se dispone de este banco de ítems, es el momento de comenzar con un proceso que, si bien no es complicado, puede resultar largo y difícil de llevar a cabo, principalmente si no se dispone de la experiencia o los conocimientos necesarios.

En el proceso general de calibración intervienen una serie de individuos que se pueden clasificar en dos tipos: Los participantes activos (responsable, supervisor, coordinador, ejecutor y otros colaboradores) – que se encargan de llevar a cabo el proceso y de poner en marcha las distintas tareas a realizar durante la calibración – y los participantes pasivos (revisores y expertos), aquéllos con los que contactarán los sujetos activos para recabar información.

Por otro lado, cualquier proceso de calibración de ítems está formado obligatoriamente por dos etapas principales: la recogida de datos de los participantes pasivos, y el análisis de esos datos y calibración de los parámetros de calibración. En conclusión, independientemente del método elegido para calibrar, se sigue siempre un procedimiento en dos fases sucesivas, que constan a su vez de una serie de actividades consecutivas a ejecutar que deberán haber sido previamente planificadas. Estas dos fases pueden dilatarse en el tiempo hasta que se haya terminado la segunda, por cumplirse un plazo máximo preestablecido como duración para el experimento o porque se haya alcanzado un volumen suficiente de datos.

En la primera fase del proceso – *Recogida de datos mediante administración* – se detalla la conducción y la recogida de los datos del experimento que posteriormente serán transcritos a una base de datos. Pudiendo emplearse multitud de técnicas para recabar información de los sujetos pasivos (Mark y Greer 1993; Arruabarrena, Pérez, Gutiérrez, López-Cuadrado y Vadillo 2002) las aportaciones de los expertos en este contexto se obtendrán mediante administración de cuestionarios a través de pruebas de campo. En la prueba de campo puede definirse un volumen mínimo de datos a recoger. Así, en esta primera fase se planifica y ejecuta la prueba de campo a través de la cual se intentará recabar el volumen de datos establecido. En esta fase de administración se incluye la recogida de valoraciones de los expertos, que se almacenarán en una base de datos.

En la segunda fase – *Análisis de datos y calibración* – se filtran los datos recogidos, se estima el parámetro dificultad de los ítems y se analiza la fiabilidad de los resultados. Esta fase parte de los datos transcritos en la base de datos y puede comenzar una vez que se cuenta con un cierto volumen de datos, es decir, que no es necesario esperar a que concluya la fase de recogida de datos.

Cada una de estas fases debe ser planificada previamente, puesto que el responsable de la calibración deberá tomar una serie de decisiones de diseño antes de su ejecución. Lo primero que se necesita especificar para poder comenzar el proceso es el conjunto de ítems a calibrar y qué tipo de calibración – expertos o psicométrica – utilizar. También se debe facilitar información que se corresponde con otras decisiones que se deben tomar en el proceso, referidas a la preparación de formularios y a las decisiones anteriores y posteriores a la consulta con los expertos. Las distintas decisiones que se deben tomar al abordar cualquier proceso de calibración se encuentran clasificadas en la Tabla 1 según su orden temporal dentro del proceso y constituyen por sí mismas una de las tareas del proceso: *diseñar el experimento a realizar*.

Tabla 1: Tipos de decisiones en el proceso integral de calibración

Tarea	1. Diseñar el experimento a realizar
1a	Especificar ítems
1b	Especificar experimento
1c	Especificar análisis de administración
1d	Especificar análisis, estudios y cálculos para calibrar

Fuente: *Elaboración propia, 2014.*

Las dos primeras especificaciones (tareas 1a y 1b de la Tabla 1) son previas a la conducción de la prueba de campo, por lo que deben realizarse antes de comenzar la ejecución de la fase de administración (Tabla 2). Concretamente, en la tarea 1a se debe *especificar el conjunto de ítems* a calibrar y en la tarea 1b se toman el resto de decisiones iniciales de planificación *del experimento*. Esta segunda tarea deja constancia de varias decisiones en el caso de una calibración vía expertos: el objetivo del experimento, los datos sobre el conjunto de expertos de que dispondrá y el diseño de los cuestionarios con los ítems a distribuir.

En cuanto a las otras dos especificaciones (tareas 1c y 1d de la Tabla 1), son decisiones previas a la estimación de los parámetros del ítem a partir de la muestra recogida. En la tarea 1c se indican los procedimientos para determinar – con independencia del resto de datos recogidos – si una administración es aceptable o no, mientras que en la tarea 1d se describen todos los procedimientos que tratan de atestiguar la fiabilidad de los ítems y de los expertos consultados una vez recogidos todos los datos, para incluirlos o no en el cálculo de la dificultad. En esta última especificación también se indican los detalles de cálculo de la dificultad a estimar.

Teniendo en cuenta esta tarea de planificación, el proceso integral de calibración consta de tres tareas principales: diseñar el experimento a realizar, ejecutar la fase de administración y ejecutar la fase de análisis y calibración (Tabla 2).

Tabla 2: Tareas en el proceso integral de calibración

<i>Tareas en la calibración de ítems</i>	<i>Requisitos</i>
Diseñar el experimento a realizar	
Ejecutar la fase de administración	1a, 1b
Ejecutar la fase de análisis y calibración	1c, 1d

Fuente: *Elaboración propia, 2014.*

De las lecciones aprendidas una vez completada la calibración del banco de ítems de Hezinet, definido el proceso de construcción de TAIs y también los procesos de calibración para los dos métodos (Arruabarrena 2010; López-Cuadrado, Pérez et al. 2010), surgió un prototipo inicial para una herramienta de ayuda a la calibración que tuviera en cuenta múltiples recomendaciones basadas en las valoraciones y conclusiones que se obtuvieron en distintos estudios donde se analizaban los costes o se realizaban comparativas (López-Cuadrado, Armendariz, Pérez y Arruabarrena 2008; López-Cuadrado, Armendariz, Pérez, Arruabarrena y Vadillo 2009). A continuación, se revisaron distintos programas software disponibles para la informatización de las distintas tareas correspondientes a ambos métodos de calibración y también se analizaron los distintos aspectos a tener en cuenta desde el momento en que se toma la decisión de recurrir a los test informatizados como mecanismo de evaluación, hasta concluir con la calibración (Presedo, Armendariz y López-Cuadrado 2012). De este modo, se fijó el objetivo de implementar una herramienta denominada CALLIE utilizando la arquitectura renovada ELSA y que permitiera al usuario calibrar un banco de ítems de cualquier materia por la vía que prefiera.

Este artículo se centra en la implementación de CALLIE y en el proceso de calibración del banco de ítems vía expertos. Uno de los objetivos, por tanto, es que la herramienta sea capaz de preparar automáticamente un sistema Web con el que autentificar a los distintos expertos participantes y administrarles los cuestionarios. Considerando que los usuarios más probables de este tipo de herramienta serán profesores, se plantea aprovechar para ese cometido una plataforma educativa Web que sea popular en ese colectivo, concretamente Moodle (gratuita y disponible en [www.moodle.org](http://www.moodle.org)), e integrarla en CALLIE. La sección 2 del artículo está dedicada a la implementación de la herramienta, la sección 3 presenta las principales aportaciones y la sección 4 muestra las conclusiones y líneas futuras de trabajo.

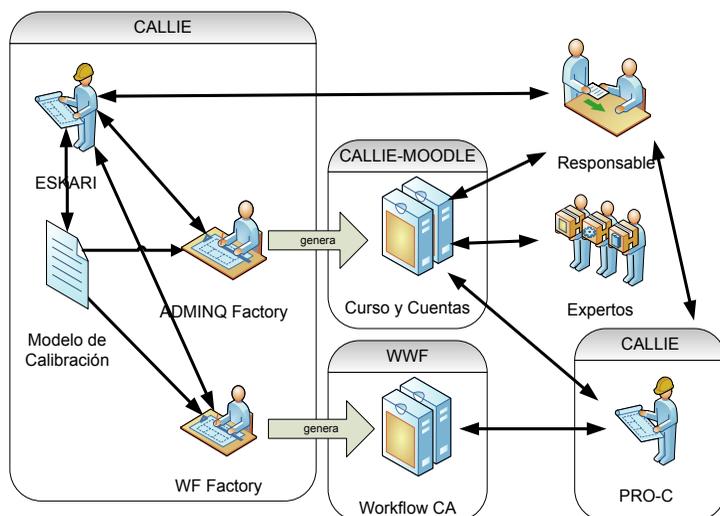
## 2. El sistema de calibración: CALLIE

El sistema de calibración se ha implementado siguiendo una arquitectura de tres capas – lógica de negocio, capa de datos e interfaz – que se detallan a continuación.

### 2.1. Lógica de negocio

La lógica de negocio del sistema se centra en la especificación del proceso de calibración a realizar, en la generación automática de otros elementos de software que permiten llevar a cabo dicho proceso y en la ejecución del propio proceso de calibración. Con este objeto, el sistema de calibración se ha dividido en tres subsistemas (Figura 1): CALLIE-MOODLE que es la plataforma educativa, WWF que es el motor de flujos de trabajo, y CALLIE que es la aplicación Web que controla el proceso integral de calibración. CALLIE se ha estructurado en los siguientes componentes: por un lado ESKARI, ADMINQ Factory y WF Factory que utilizan el modelo de calibración como elemento de comunicación entre ellos, y por otro PRO-C que se comunica con los elementos software creados en los otros dos subsistemas.

Figura 1: Arquitectura del sistema de calibración



Fuente: Elaboración propia, 2014.

En cuanto a los usuarios, el *responsable* aglutina todos roles de los participantes activos del proceso y dispondrá de acceso a las interfaces de los distintos módulos de CALLIE. También, una vez registrado en CALLIE tendrá acceso automático a un curso de CALLIE-MOODLE como profesor cuando solicite una calibración mediante el juicio de expertos. Los sujetos *expertos*, por su parte, tendrán la labor de rellenar los cuestionarios que establezca el propio sistema CALLIE pero utilizando la plataforma Moodle. Por ello, un experto no necesitará de ningún rol en el sistema CALLIE, solo necesitará tener acceso como alumno al curso generado en CALLIE-MOODLE.

*ESKARI* es el primer componente con el que interactúa el responsable de la calibración. Se encarga de la tarea de diseño del experimento obteniendo la especificación del proceso de calibración que se quiere llevar a cabo (*modelo de calibración*). También se encarga de solicitar a ADMINQ Factory y a WF Factory la creación de los artefactos necesarios para la ejecución de las fases de administración y de análisis de datos y calibración respectivamente.

*ADMINQ Factory* es el componente que, partiendo del modelo de experimento especificado mediante ESKARI, se encarga de crear automáticamente un artefacto software para la administra-

ción electrónica mediante Web de los ítems especificados en una determinada calibración: un *curso en la plataforma Moodle*, junto con las *cuentas de acceso al curso* de los distintos agentes participantes. Para que el curso generado sea compatible con CALLIE-MOODLE utiliza un estándar: el cartucho común o IMS CC (Riley y Mills 2008), que representa un *curso reutilizable* y que contiene, representados en IMS QTI, los distintos cuestionarios a administrar y los ítems a calibrar. Con cada curso se da servicio a dos tipos de usuarios, los expertos que van a responder a los ítems y el responsable del experimento. Posibilita la tarea de ejecución de la fase de administración con el apoyo de CALLIE-MOODLE.

*CALLIE-MOODLE* es la plataforma Moodle integrada en el sistema de calibración para realizar la parte de administración de los cuestionarios. Este Moodle se encuentra por defecto en el mismo servidor que CALLIE y se configura automáticamente para acomodarse a las necesidades de administración a expertos. Acoge todos los módulos generados automáticamente mediante ADMINQ Factory, es decir, cada curso correspondiente a cada calibración junto con las cuentas de acceso.

*WF Factory* es el componente que, partiendo de los cálculos y análisis explicitados en el modelo de experimento especificado mediante ESKARI, genera un flujo de trabajo denominado *workflow CA*. El *workflow CA* necesita del motor de Microsoft Windows Workflow Foundation (*WWF*) para ejecutarse y representa, en un orden concreto, las tareas de análisis de administraciones, los cálculos necesarios para la estimación del parámetro dificultad y los estudios, generalmente estadísticos, que permiten sentar las bases de una calibración de calidad. Para su creación, el sistema se sirve de una *biblioteca* que contiene varios criterios, estudios y análisis posibles para utilizar en calibraciones de ítems mediante el juicio de expertos. El *workflow CA* consta de dos flujos consecutivos: uno que realizará el análisis de administraciones y otro que realizará los análisis de fiabilidad y calibración. Este *workflow CA* posibilita, por lo tanto, la tarea de ejecución de la fase de análisis de datos y calibración.

*PRO-C* es el otro componente de CALLIE con el que interactúa el responsable. Se encarga de llevar el control del proceso de calibración una vez generados los artefactos invocando en un orden concreto la ejecución de las distintas tareas que lo forman, para lo cual se comunica tanto con el curso generado en Moodle como con el *workflow CA* y también con el responsable, lo que permite a este último controlar y monitorizar sus calibraciones en curso.

## 2.2. La capa de datos

Respecto a la capa de datos, la realización de cada una de las tres tareas del proceso da lugar a un modelo de información con los datos necesarios para acometer su función adecuadamente: el modelo de calibración, el modelo de administración de cuestionarios, y el modelo de análisis y cálculos a partir de la administración de cuestionarios (ADMINQCA). El modelo ADMINQCA aglutina los datos recogidos y relevantes para la realización de los distintos análisis y cálculos junto con los resultados de esos análisis y cálculos. Los tres modelos mencionados constituyen la capa de datos de la arquitectura, de manera que CALLIE-MOODLE almacena en la base de datos de la plataforma Moodle – denominada *moodle* e implementada en MySQL – todos los datos correspondientes al modelo de administración de cuestionarios. Por otro lado, la base de datos propia de la aplicación – denominada *CALLIE\_DB* e implementada en MS SQL Server 2008 – almacena todos los datos correspondientes a los demás modelos, incluidos los de las especificaciones e ítems mediante modelos XML.

## 2.3. La interfaz

En cuanto a la *interfaz* de CALLIE, como la audiencia esperada de la herramienta son profesores de instituto, de ciencias o de idiomas, y no psicómetras ni informáticos ni expertos en el tema de calibración, es una interfaz informal, sencilla y flexible que está convenientemente preparada para ayudar a personas involucradas en sistemas de aprendizaje con poca o ninguna experiencia en calibración. Surge como consecuencia del estudio de la comunicación utilizando un modelo de género del sistema de aplicación de test adaptativos informatizados (López-Cuadrado, Armendariz, Latapy y Lopistéguy 2008). Como la intención de la herramienta es guiar al usuario inexperto a realizar sus

calibraciones con éxito, la aplicación ofrece varios tipos de ayuda y previene de realizar alguna acción cuando no está permitida por la lógica de negocio (ver Figura 2). Para comenzar, ofrece ayuda contextual sobre cualquier objeto de la pantalla al situar el ratón sobre el elemento en cuestión. La herramienta CALLIE utiliza un código de colores (rojo, naranja y verde) que acompaña con los avisos textuales oportunos para que el usuario visualice rápidamente las incidencias que ocurran. En la interfaz también aparecen valores sombreados que representan valores predefinidos no aplicables al caso actual o no modificables (fondo gris) o, si sí son modificables, que no se pueden modificar en ese paso, generalmente porque se han decidido con anterioridad (fondo azul). A la hora de hacer las especificaciones CALLIE ofrece guía adaptativa porque siempre ofrece una opción válida según el criterio del sistema para la decisión que se debe tomar y adaptándose a las decisiones que se tienen hasta el momento. El sistema, además, vigila que se han realizado todos los componentes de la decisión y detecta si nuevas decisiones afectan a las ya tomadas y, en tal caso, advierte de los efectos que suponen las mismas. También deja constancia, mediante el código de colores, que hay algún problema asociado a algún valor. En caso de problemas, siempre se puede restablecer la opción válida sugerida por el sistema. Por último, existe la posibilidad de probar el sistema con el banco EUSK, el banco de ítems utilizado para Hezinet.

Figura 2: Página de CALLIE mostrando varias características de la interfaz

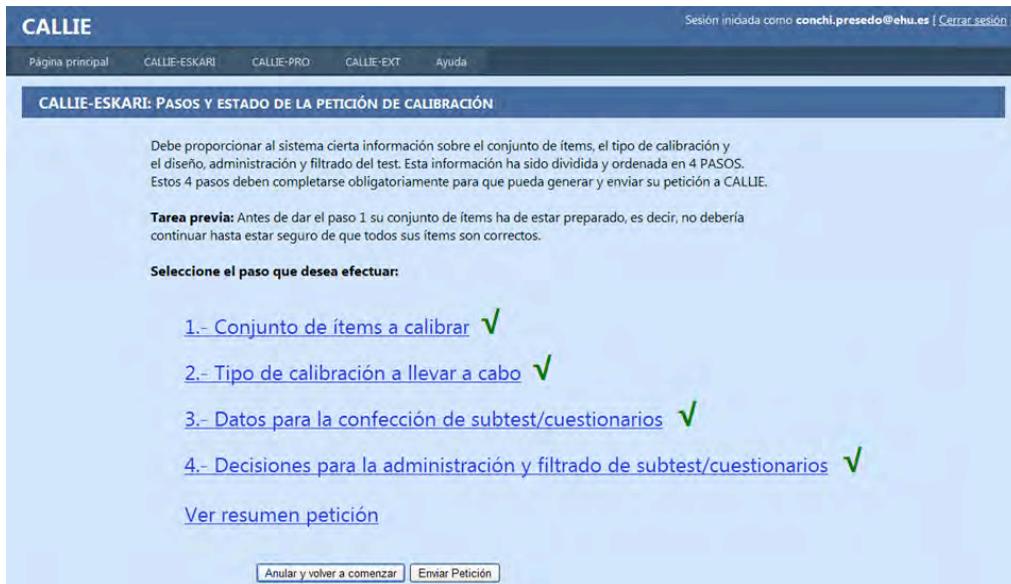
Fuente: Elaboración propia, 2014.

Por otro lado, el proceso de calibración se implementa, desde el punto de vista de la información que debe aportar al mismo el responsable de la calibración, mediante un proceso en el que el responsable registra y envía al sistema una *petición de calibración* con toda la información necesaria para el proceso, el sistema procesa la calibración solicitada y muestra el resultado final de la misma al solicitante para obtener su aprobación o su rechazo. Así, el proceso de calibración se automatiza mediante un workflow específico que depende del tipo de calibración y que consta de dos subprocesos consecutivos: *Registrar y enviar petición de calibración* y *Procesar petición de calibración*. En el primero de ellos el responsable rellena los datos oportunos y envía la solicitud de calibración y en el segundo de ellos se ejecutan las distintas tareas del proceso de calibración utilizando esos datos. En el momento en que el responsable envía la petición de calibración se crea el *modelo de calibración* que refleja todas las decisiones tomadas. Asociada a cada petición existe un proceso de calibración dinámico que puede estar en fase de especificación, administración, calibrando o finalizado. Estas fases generan en cada momento del proceso un estado para la calibración, que pasará sucesivamente por los estados *iniciada* (se está especificando), *recibida* (se ha enviado), *aceptada* (se ha validado el modelo de calibración), *administrada* (se han recogido los datos de las

administraciones a los expertos), *calibrada* (se han realizado los análisis y estimado el parámetro dificultad) o *finalizada* (se han aceptado o rechazado los resultados). Además, en cualquier momento antes de su finalización, el proceso se puede cancelar y la calibración pasará a estado *abortada*. Cada responsable puede especificar varias peticiones de calibración.

Por ello, la *interfaz* de CALLIE posee dos partes bien diferenciadas, que se encuentran en los componentes de la arquitectura con el mismo sufijo, en forma de dos submenús principales que permiten al responsable planificar una o varias calibraciones y controlar su ejecución: CALLIE-ESKARI (Figura 3) y CALLIE-PRO.

Figura 3: Menú de pasos de CALLIE-ESKARI una vez dados los cuatro pasos

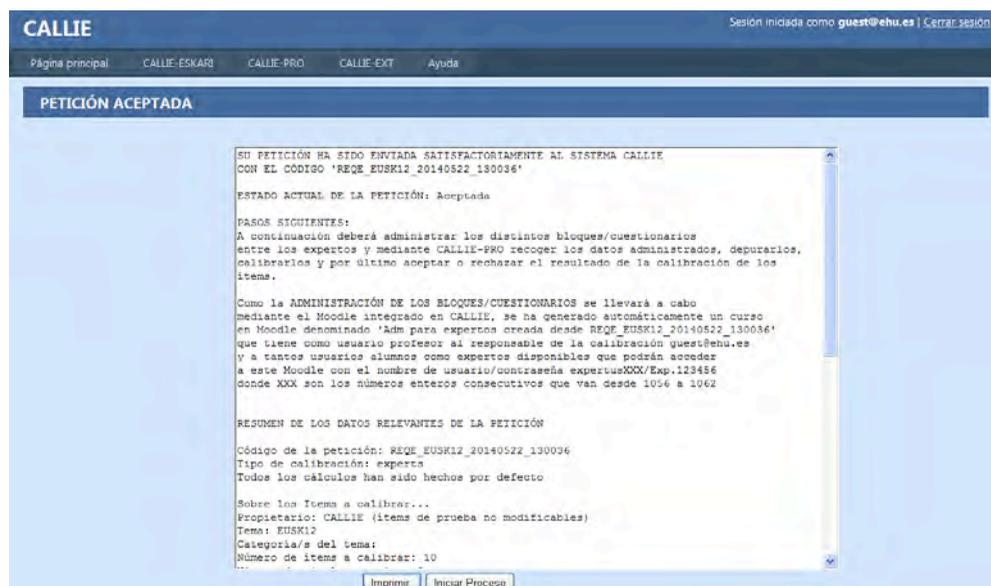


Fuente: Elaboración propia, 2014.

Mediante *CALLIE-ESKARI* el responsable de la calibración realiza una especificación inicial y la envía al sistema (subproceso registrar y enviar petición de calibración). Este submenú solicita los datos que se precisan para poder llevar a cabo un experimento de calibración vía expertos a través de varias páginas Web cuyo orden de navegación coincide básicamente con el de las decisiones mencionadas en la Tabla 1. Primero se establece cuál es el banco de ítems que se quiere calibrar y qué tipo de calibración se desea efectuar. Seguidamente, se diseñan los cuestionarios para administrar a los expertos y se toman otras decisiones, sobre los datos a pedir por ítem más otras referidas al tratamiento automático que el sistema dará a los datos recogidos de los expertos. Así, *CALLIE-ESKARI* se ha dividido en los cuatro pasos que aparecen en la Figura 3 y que *CALLIE* recomienda dar en orden.

Cuando el usuario pulsa *Enviar Petición*, que solamente se habilita en *CALLIE-ESKARI* después de realizar los cuatro pasos de su menú, la petición pasa del estado iniciada al estado recibida. En ese momento *ESKARI* crea el modelo de calibración XML correspondiente a la petición, y lo valida contra su esquema XSD. Si no es correcto se avisa al usuario y se cancela el proceso, con lo que la petición pasará a abortada. Si *CALLIE* da por válida la petición, se le asigna un código único y en función de él, el sistema sabrá a qué tipo de calibración corresponde: psicométrica (si comienza con REQP) o vía expertos (si comienza con REQE), pasará a aceptada, y se redirige al usuario a una pantalla similar a la de la Figura 4 para confirmar que desea continuar.

Figura 4: Página para iniciar el proceso de calibración



Fuente: *Elaboración propia, 2014.*

En esta pantalla se avisa al responsable de que su petición ha sido aceptada, de los siguientes pasos que deberá seguir y de las decisiones relevantes tomadas mediante un resumen. Cuando el usuario confirme que desea continuar, se inicia el proceso correspondiente creando los artefactos software necesarios. A partir de ese momento, el control de su progreso hasta su finalización o cancelación, así como las posibles modificaciones sobre la petición inicial (subproceso procesar petición de calibración) pueden ser gestionadas por parte del responsable exclusivamente a través CALLIE-PRO.

Mediante el submenú *CALLIE-PRO* el responsable podrá controlar y monitorizar el progreso y cambiar las especificaciones de sus calibraciones. En todo momento es guiado por el sistema en estas tareas, de modo que se le ofrecen exclusivamente las opciones permitidas en cada momento – que dependen del estado actual y del tipo de administración de la petición seleccionada – y se le oculta el resto. Para evitar posibles interferencias entre pasos, lo único que se le permitirá modificar son los parámetros de filtrado de administraciones e ítems. En cualquier otro caso la única opción será abortar el experimento en curso y planificar una nueva petición. Independientemente del tipo de calibración a realizar, el responsable se comunica con la aplicación a través de dos páginas Web, que se presentan a continuación.

En la primera de ellas – *Procesar mis peticiones de calibración* (Figura 5) – el usuario puede ver una lista con todas las peticiones que ha solicitado, establece sobre cuál de ellas quiere operar y CALLIE le presenta las distintas acciones que puede efectuar. Se han utilizado colores para indicar la finalización con éxito (verde) o no (rojo) de la calibración.

Figura 5: Página Procesar peticiones seleccionando una petición “Calibrada”

**CALLIE-PRO.- PROCESAR MIS PETICIONES DE CALIBRACIÓN**

Calibraciones en curso del usuario: [guest@ehu.es](#)

Mediante esta página podrá visualizar y/o realizar cualquiera de las operaciones permitidas para la petición que elija y que dependan del estado de la misma. Las calibraciones terminadas se indican mediante su estado en color rojo o verde (terminada con éxito). En naranja aparecen los códigos de las peticiones que han superado el plazo máximo de 180 días para completar el proceso de calibración.

**Selección de la siguiente lista la petición en curso a procesar:**

	código de la petición	fecha y hora petición	tipo de calibración	banco de ítems	tipo de administración	estado	fecha y hora estado
<a href="#">Seleccionar</a>	<a href="#">REQE_kk100_20130529_215150</a>	29/05/2013 21:51:53	vía expertos	bancoguest2	Moodle de CALLIE	Abortada (ACE)	23/04/2014 11:57:59
<a href="#">Seleccionar</a>	<a href="#">REQE_kk100_20130529_211404</a>	29/05/2013 21:14:50	psicométrica	bancoguest2	Otra	Aceptada	01/06/2013 9:17:57
<a href="#">Seleccionar</a>	<a href="#">REQE_EUSK12_20140414_232816</a>	14/04/2014 23:35:25	vía expertos	EUSK12	Moodle de CALLIE	Aceptada	14/04/2014 23:35:25
<a href="#">Seleccionar</a>	<a href="#">REQE_kk1_20130227_134759</a>	27/02/2013 13:48:07	vía expertos	bancoguest1	Moodle de CALLIE	Calibrada	23/05/2013 12:00:16
<a href="#">Seleccionar</a>	<a href="#">REQE_EUSK12_20130601_121618</a>	01/06/2013 12:17:27	vía expertos	EUSK12	Moodle de CALLIE	Calibrada	22/05/2014 11:41:01
<a href="#">Seleccionar</a>	<a href="#">REQP_kk1_20130226_080745</a>	26/02/2013 8:07:49	psicométrica	bancoguest1	Otra	Finalizada	01/06/2013 9:17:57

[Eliminar/Abortar](#)
[Ir al Moodle de CALLIE](#)
[Ver Resumen](#)
[Ver resultados](#)

Fuente: Elaboración propia, 2014.

En la segunda página de CALLIE-PRO – *Ver resultado de la calibración* (Figura 6) – el responsable puede visualizar los resultados de la calibración y decidir si los acepta o no. Se puede acceder a ella una vez que la petición alcanza el estado de calibrada (mediante el botón *Ver resultados* de la Figura 5) y ofrece un *resumen con los resultados finales* divididos en tres partes: parámetros utilizados en los análisis, resumen de ítems y resumen de administraciones a expertos.

Figura 6: Página de CALLIE Ver resultados de la calibración

**CALLIE-PRO.- VER RESULTADO DE LA CALIBRACIÓN**

Resultados de la calibración con código: [REQE\\_EUSK12\\_20130601\\_121618](#)

**Parámetros utilizados**

10 ítems en el banco y 10 niveles posibles de dificultad. Se permite eliminar ítems (respuestas correctas < 70% y menos del 85% en una horquilla de 3 niveles consecutivos). Se permite rechazar administraciones (incorrectas).

**Resumen ítems**

cod ítem	título	dificultad	estado	razón del estado
271 <a href="#">EUSK1</a>	Zuek zer zaretze?	2	aviso	3 juicios válidos: Se recomienda recoger un mínimo de 7.
272 <a href="#">EUSK2</a>	Norena da libburua?		eliminado C.it-1:	El porcentaje de respuestas correctas (33,33%) no llega al 70%
273 <a href="#">EUSK3</a>	Non dago Urgull mendia?	2,6667	aviso	3 juicios válidos: Se recomienda recoger un mínimo de 7.
274 <a href="#">EUSK4</a>	Gu orain Donostian gaude.		eliminado C.it-1:	El porcentaje de respuestas correctas (66,67%) no llega al 70%
275 <a href="#">EUSK5</a>	Urgull mendia handia da.		eliminado C.it-1:	El porcentaje de respuestas correctas (66,67%) no llega al 70%

[1 2](#)

**Resumen administraciones expertos**

nº cuestionario/bloque	idexperto	estado	razón del estado
1	<a href="#">conchi_presedo@ehu.es</a>	rechazado	C.ex-1: No estima adecuadamente el nivel en ningún ítem
2	<a href="#">conchi_presedo@ehu.es</a>	rechazado	C.ex-1: No estima adecuadamente el nivel en ningún ítem
1	<a href="#">expertus131</a>	aceptado	
1	<a href="#">expertus132</a>	aceptado	
1	<a href="#">expertus133</a>	aceptado	

[1 2](#)

[<< Volver](#)
[Descargar detallado](#)
[Recoger más datos](#)
[Cambiar filtros y simular](#)
[Rechazar resultados](#)
[Aceptar resultados](#)

Fuente: Elaboración propia, 2014.

Para facilitar la revisión de estos resultados, además de avisos textuales, se ha establecido el código de colores para indicar aceptación (verde), aceptación aunque no cumple las recomendaciones (naranja) y rechazo o eliminación (rojo). Por último, CALLIE es capaz de generar un documento Excel con los cálculos detallados realizados por la herramienta, recoger más datos desde Moodle o desde Excel en función del tipo de administración seguido, y cambiar los filtros y simular nuevos resultados de calibración modificando los parámetros de análisis sobre los datos recogidos.

Además, en toda petición no abortada y cuyo tipo de administración sea mediante el Moodle del sistema de calibración, el botón *Ir al Moodle de CALLIE* redirige al responsable a la página principal de CALLIE-MOODLE (Figura 7) en la que podrá acceder a los cursos asociados a sus peticiones y controlar sus administraciones, cuestionarios y expertos utilizando Moodle. Las flechas de navegación del navegador permiten volver a CALLIE.

Figura 7: Página principal de CALLIE-MOODLE



Fuente: *Elaboración propia, 2014.*

### 3. Resultados

En primer lugar se ha analizado el proceso global de calibración de un banco de ítems independientemente del método seguido y se ha obtenido el conocimiento necesario para establecer una base común sobre la que poder realizar el diseño, control y seguimiento de cualquier proceso de calibración. Con este conocimiento, se ha diseñado e implementado el sistema CALLIE, capaz de guiar al usuario inexperto a lo largo de todo el proceso de calibración, sin que sea necesario que éste cuente con conocimientos previos en el área. El sistema utiliza técnicas de desarrollo software dirigido por modelos, concretamente la generación automática de software, y se basa en la tecnología de workflows para la creación de actividades de análisis y cálculos que se pueden organizar de diferentes modos. La herramienta implementada dispone además de un banco de ítems de prueba con el que el usuario puede efectuar simulaciones de calibración con distinta cantidad de ítems, tanto basadas en el juicio de expertos como fundamentadas en la TRI. También dispone de un script que genera automáticamente el modelo de datos común, CALLIE\_DB, lo que permite que la herramienta pueda adaptarse fácilmente a distintos sistemas de gestión de bases de datos.

En segundo lugar se ha analizado el proceso de calibración de un banco de ítems mediante el juicio de expertos y se ha obtenido el conocimiento necesario para poder especificar el experimento de administración de ítems que conlleva, concretado en las decisiones que se tienen que tomar y cómo interfieren entre ellas. De este conocimiento ha surgido ESKARI, el componente de especificación del experimento de calibración. Se ha generado un modelo XML que ha sido unificado con el correspondiente a la calibración de tipo psicométrico para garantizar la integración de los dos métodos en un único sistema.

Se han estudiado los métodos de aplicación de un cuestionario a un experto y se ha creado ADMINQ Factory, un módulo que, utilizando las técnicas de desarrollo dirigido por modelos y una plataforma educativa preexistente que permita la importación de paquetes IMS, genera automáticamente un sistema de administración de cuestionarios a expertos basado en web. Esto supone que el sistema CALLIE puede aprovechar todas las funcionalidades que posea la propia plataforma educativa en la que se crea. Para ello se ha propuesto un nuevo modelo de ítem a utilizar en las calibraciones mediante expertos, que surge de los ítems de evaluación más comunes. Este modelo de ítem ha sido implementado en IMS QTI lo que posibilita su importación directa a cualquier sistema que contemple ese estándar con independencia de la versión utilizada. Partiendo de este nuevo modelo de ítem y de la especificación de la calibración se ha creado un programa software capaz de obtener el paquete IMS CC con el que posteriormente se genera un curso en Moodle para administrar los cuestionarios a los expertos disponibles. Por último, se ha ideado un servicio Web para Moodle que es capaz de importar paquetes IMS CC creando automáticamente el curso de administración y se han detectado servicios ya existentes en la plataforma para el alta y matriculación automática de expertos en el curso previamente generado. Se ha preparado el modelo de datos para su soporte.

La Figura 8 muestra un ítem para ser calibrado vía expertos generado automáticamente por CALLIE en el curso de Moodle tal y como lo vería el experto #4 que está identificado en el sistema. CALLIE siempre genera tres campos para cada ítem a calibrar, el primero corresponde al propio ítem que el experto deberá contestar. En el segundo campo el experto estimará el nivel de dificultad para ese ítem, en este caso mediante un valor comprendido entre 1 y 11, y en el tercero podrá especificar cualquier otro dato correspondiente al ítem si así se lo indica el responsable.

Figura 8: Ítem generado automáticamente en CALLIE-MOODLE

The screenshot shows a Moodle interface for 'EUSK IRALE Expert-Based Calibration 6'. It features a navigation menu on the left with buttons for questions 1 through 12. The main content area displays three questions:

- 4. galdera:** 'Kaixo, Aimar! Azkenean abenduan gaudu, eta astelehenean ospatuko genuen zure bosgarren urtebetetzea. Zelako jurgal! Amatzu despietatuta ibili da, baina hemen dago zure agurra. Zorionak, mutil handi, Maialenen, aitaren eta amatzoren partetik. Noiz jaiotakoa da Aimar?' with radio button options: 'a. Urtearen hasieran.', 'b. Urtearen erdi aldera.', 'c. Gabonetatik gertu.', 'd. Aste Santuan.' and an 'Egiaztatu' button.
- 5. galdera:** 'Zailtasuna (1-tik 11-ra):' with an 'Erantzuna:' input field and an 'Egiaztatu' button.
- 6. galdera:** 'Besteak:' with an 'Erantzuna:' input field and an 'Egiaztatu' button.

At the bottom, there is a 'Hurrengoa' button and a footer with 'Expert User4 barruan da (Itzen)' and 'EUSKIRAILEXPERT-BASEDCALIBRATION6'.

Fuente: *Elaboración propia, 2014.*

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 – proporciona múltiples ventajas, entre ellas permite la creación de contenidos reutilizables y facilita la interoperabilidad entre distintos sistemas (Fallon y Brown 2003).

Para finalizar, el sistema genera automáticamente en formato Excel una serie de informes predefinidos con los cálculos detallados y resultados concernientes al proceso de calibración, que pueden ser objeto de análisis posteriores con otros programas estadísticos que soporten la importación de datos desde Excel, como por ejemplo R. El sistema también es capaz de capturar los datos de las administraciones a expertos desde un fichero Excel con un formato predefinido.

#### 4. Conclusiones y trabajo futuro

Existe una serie de líneas abiertas para mejorar y dar continuidad a la labor realizada.

La primera línea futura consistiría en mejorar la herramienta solucionando las debilidades detectadas, a saber: actualmente varias decisiones se toman basándose en criterios fijos que son consecuencia de la experiencia, los métodos implementados para realizar automáticamente el diseño de cuestionarios de los bancos de ítems son rudimentarios, CALLIE solo permite la introducción de ítems individuales de respuesta múltiple, con un enunciado y varias posibles respuestas textuales, el sistema implementa exclusivamente los análisis de administraciones, estudios y método de calibración que utilizó Arruabarrena (2010) y la importación/exportación con Excel se realiza a nivel muy básico. Con objeto de mejorar la herramienta, se podrían revisar todas las decisiones de ESKARI y parametrizar al máximo los procesos de decisión; mejorar los métodos de diseño de cuestionarios incluyendo una serie de criterios más avanzados para realizar el reparto de los ítems; incluir nuevos formatos de ítems en el sistema que ya contempla el estándar IMS QTI, como localizar un punto en un dibujo o foto, unir puntos para formar una figura, presentación de videos o música, o una sopa de letras o incluir nuevos métodos de especificación de los bancos de ítems; incluir métodos alternativos para estimar la dificultad, nuevas propuestas para el análisis de administraciones e ítems, e incluso otro tipo de estudios.

Otra posible vía de mejora sería ampliar CALLIE para que el sistema fuese capaz de crear informes de forma automática conocidas las características que se quieren analizar. También podría interesar la generación de estadísticas demográficas e informes sobre los expertos que han participado en la calibración.

Aunque en la actualidad CALLIE permite la importación de los datos de las administraciones en formato Excel, en el caso de un sistema combinado, sería útil la inclusión de un módulo que permita incorporar las que se han hecho manualmente, empleando aparatos de lectura óptica para capturar las respuestas a un cuestionario.

El programa software que utiliza CALLIE para generar cartuchos IMS CC con el curso de administración para calibración vía expertos también está preparado para generar paquetes de contenido IMS CP (Smythe y Nielsen 2007) con el curso, que es otro método de distribuir contenidos entre sistemas. Se podría explorar la posibilidad de administrar los ítems a los expertos mediante otras plataformas educativas diferentes a Moodle que admitan cualquiera de estos dos estándares, como por ejemplo Atutor, Claroline, Ilias, o Sakai.

Actualmente, se ha creado un módulo para Smartphones que facilita la gestión de expertos en las administraciones. Esta versión para móviles permite al responsable comunicarse con Moodle creando una cuenta para un nuevo experto, matriculándolo en el curso que se le indique como alumno y/o eliminando a un experto previamente creado, utilizando los servicios web del propio Moodle. En un futuro se prevé completar este módulo de forma que soporte el resto de funcionalidades de CALLIE, pero a través del teléfono móvil.

Por último, también se está considerando la aplicación de otras teorías de calibración. El sistema se basa en la actualidad en ítems independientes, sin embargo, en algunas ocasiones es necesario que algunos ítems se presenten al usuario en cierto orden o que vayan ligados entre sí. Para modelar este tipo de ítems o conjunto de ítems existe la teoría de respuesta al testlet (Wainer, Bradlow y Du 2000), que no se contempla en CALLIE. Es una línea abierta que puede tener muchas aplicaciones, una de ellas es la evaluación del comportamiento en casos clínicos para estudiantes de medicina, en la que se está trabajando actualmente.

## REFERENCIAS

- Armendariz, A. J., López-Cuadrado, J., Tapias, A., Villamañe, M., Sanz-Lumbier, S. y Sanz-Santamaría, S. (2003). *Learning environments should follow standards: ELSA does*. World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education, Phoenix, Arizona (USA), Association for the Advancement of Computing in Education (AACE).
- 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*. San Sebastián, España: Lenguajes y Sistemas Informáticos, Universidad del País Vasco/Euskal Herriko Unibertsitatea.
- Arruabarrena, R., Pérez, T. A., Gutiérrez, J., López-Cuadrado, J. y Vadillo, J. A. (2002). *On Evaluating Adaptive Systems for Education*. AH2002, 2nd. International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, Málaga, España: Springer-Verlag.
- Fallon, C. y Brown, S. (2003). *E-learning Standards: A guide to purchasing, developing and deploying standards-conformant e-learning*. Ed. ST. Lucie Press.
- Kingsbury, G. G. y Weiss, D. J. (1983). A comparison of IRT-based Adaptive Mastery Testing and Sequential Mastery Testing Procedure. En D. J. Weiss (eds.), *New Horizons in Testing: Latent Trait Test Theory and Computerized Adaptive Testing* (pp. 257-283). New York, United States: Academic Press.
- López-Cuadrado, J. (2010). *Test adaptativos informatizados de ingreso en un sistema e-learning*. Saarbrücken, Germany: Lambert Academic Publishing.
- López-Cuadrado, J., Armendariz, A. J., Latapy, M. y Lopistéguy, P. (2008). A genre-based perspective for the development of communicative computerized adaptive tests. *Journal of Educational Technology & Society*, 11(1), pp. 87-101.
- López-Cuadrado, J., Armendariz, A. J., Pérez, T. A. y Arruabarrena, R. (2008). *Helping tools for item bank calibration and development of computerized adaptive tests*. International Technology, Education, and Development Conference (INTED2008). Valencia, España: International Association of Technology, Education, and Development.
- López-Cuadrado, J., Armendariz, A. J., Pérez, T. A., Arruabarrena, R. y Vadillo, J. A. (2009). Computerized adaptive testing, the item bank calibration and a tool for easing the process. En A. Lazinec & C. Calafate (Ed.), *Technology Education and Development* (pp. 457-478). New York, United States: Intech.
- 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), pp. 1044-1055.
- Mark, M. A. y Greer, J. E. (1993). Evaluation Methodologies for Intelligent Tutoring Systems. *International Journal of Artificial Intelligence in Education. Special Issue on Evaluation*, 4(2/3), pp. 129-153.
- Muñiz, J. (2010). Las teorías de los tests: Teoría Clásica y Teoría de Respuesta a los Ítems. *Papeles del Psicólogo*, 31, pp. 57-66.
- Pérez, T. A. (2000). *Un hiperentorno adaptativo para el aprendizaje instructivo / constructivo*. San Sebastián, España: Lenguajes y Sistemas Informáticos, Univ. País Vasco/Euskal Herriko Unibertsitatea.
- 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, Germany: Editorial Académica Española.
- Riley, K. y Mills, D. (2008). *IMS Common Cartridge (CC) Authorization Web Service v1.0 Final Specification*. IMS Global Learning Consortium, Inc., accedido el 10/6/2104 desde [www.imsglobal.org/cc/](http://www.imsglobal.org/cc/)
- Smythe, C. y Nielsen, B. (2007). *IMS Content Packaging (CP) Specification Primer v1.2 Public Draft*. IMS Global Learning Consortium, Inc., accedido el 10/6/2014 desde [www.imsglobal.org/content/packaging/](http://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*. IMS Global Learning Consortium, Inc., accedido el 10/6/2014 desde [www.imsglobal.org/question/](http://www.imsglobal.org/question/)
- Wainer, H., Bradlow, E. T. y Du, Z. (2000). Testlet response theory: an analog for the 3PL model useful in testlet-based adaptive testing. En W. J. van der Linden & C. A. W. Glas (Ed.), *Computerized adaptive testing: theory and practice* (pp. 245-269). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F., Mislevy, R. J., Steinberg, L. y Thissen, D. M. (2000). Future challenges. *Computerized Adaptive Testing: A Primer (2nd edition)* (pp. 231-269). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

## SOBRE LOS AUTORES

**Concepción Presedo:** Profesora del departamento Lenguajes y Sistemas informáticos. Actualmente imparte clase en segundo curso del Grado de Ingeniería Informática de Gestión y Sistemas Informáticos y primer curso del Grado de Ingeniería Industrial en la Escuela Universitaria de Ingeniería Técnica Industrial de Bilbao. Su investigación se centra en informática educativa, concretamente en la calibración de ítems para su posterior utilización en test adaptativos informatizados. Ha sido partícipe del desarrollo de un sistema que ayuda a usuarios poco expertos en estadística, informática y psicometría a llevar a cabo calibraciones de bancos de ítems, para ello se han utilizado métodos de desarrollo dirigidos por modelos y workflows.

**Anaje Armendariz:** Profesora del departamento Lenguajes y Sistemas informáticos. Actualmente imparte clases en segundo y tercer curso del Grado de Ingeniería Informática de Gestión y Sistemas Informáticos en la Escuela Universitaria de Ingeniería Técnica Industrial de Bilbao. Su investigación se centra en informática educativa, concretamente en la calibración de ítems para su posterior utilización en test adaptativos informatizados. Ha sido partícipe del desarrollo de un sistema que ayuda a usuarios poco expertos en estadística, informática y psicometría a llevar a cabo calibraciones de bancos de ítems, para ello se han utilizado métodos de desarrollo dirigidos por modelos y workflows.

**Javier López-Cuadrado:** Profesor doctor del departamento Lenguajes y Sistemas informáticos. Actualmente imparte clase en primero del Grado de Ingeniería Informática de Gestión y Sistemas Informáticos en la Escuela Universitaria de Ingeniería Técnica Industrial de Bilbao. Su investigación se centra en informática educativa, concretamente en la calibración de ítems para su posterior utilización en test adaptativos informatizados. Ha dirigido el desarrollo de un sistema que ayuda a usuarios poco expertos en estadística, informática y psicometría a llevar a cabo calibraciones de bancos de ítems, para ello se han utilizado métodos de desarrollo dirigidos por modelos y workflows.

**Tomás A. Pérez:** Profesor doctor del departamento Lenguajes y Sistemas informáticos. Actualmente imparte clase en segundo y tercero del Grado de Ingeniería Informática de la Facultad de Informática de San Sebastián. Su investigación se centra en informática educativa, concretamente en la calibración de ítems para su posterior utilización en test adaptativos informatizados. Ha dirigido el desarrollo de un sistema que ayuda a usuarios poco expertos en estadística, informática y psicometría a llevar a cabo calibraciones de bancos de ítems, para ello se han utilizado métodos de desarrollo dirigidos por modelos y workflows.