

Una herramienta de diseño y análisis de instrumentos de evaluación e indagación docente

Ana Isabel Molina¹, Ángel Luis Wizner¹, Carmen Lacave¹, Jesús Gallardo²

¹ Dpto. de Tecnologías y Sistemas de la Información, Ciudad Real, Universidad de Castilla-La Mancha

² Dpto. de Informática e Ing. Sistemas, Teruel, Universidad de Zaragoza

AnaIsabel.Molina@uclm.es, AngelLuis.Wizner@alu.uclm.es,
Carmen.Lacave@uclm.es, Jesus.gallardo@unizar.es

Resumen

Es habitual que los profesores hagan uso de diversos instrumentos de evaluación e indagación durante su labor docente e investigadora, tales como cuestionarios, escalas de opinión y actitudes o pruebas de conocimiento tipo test. Existe una gran variedad de sitios web que ofrecen funcionalidades para la creación y aplicación de encuestas. Sin embargo, el análisis de resultados que realizan estas herramientas es, en la mayoría de los casos, limitado. Con el objetivo de cubrir algunas de las necesidades detectadas, se ha desarrollado una aplicación web para la gestión y diseño de cuestionarios de opinión y test de conocimientos. Dicha aplicación considera aspectos de usabilidad y accesibilidad web, e incluye un repositorio de recomendaciones y *checklist* para el diseño de cuestionarios y pruebas cognitivas. La aplicación incorpora un completo módulo de análisis de resultados, en el que se calculan gran cantidad de propiedades métricas de los instrumentos de evaluación diseñados.

Abstract

It is common for teachers to make use of several assessment and inquiry instruments during his teaching and research activity, such as questionnaires, opinion and attitudes scales or knowledge tests. There are a variety of websites that offer functionalities for creating and conducting surveys. However, the analysis of results performed by these tools is, in most cases, limited. With the aim of covering some of the needs identified, a web application has been developed for the management and design of opinion surveys and knowledge tests. This application considers aspects of web usability and accessibility, and includes a repository of guidelines and checklist for designing questionnaires and cognitive tests. The application incorporates a complete module for analysis of results, in which a large numbers of metric properties of the assessment instruments designed are calculated.

Palabras clave

Evaluación de instrumentos de medida, fiabilidad, encuestas, pruebas cognitivas, recurso docente, métricas, calidad, evaluación de competencias.

1. Introducción

La creciente demanda por parte empresas e instituciones públicas y privadas de mecanismos que ayuden a conocer el grado de satisfacción de sus clientes y/o usuarios, hacen necesaria la aparición de aplicaciones para conocer su opinión acerca de los servicios prestados o productos disponibles.

El ámbito académico no queda exento de estas necesidades. Una universidad necesita conocer sus puntos fuertes y débiles en aspectos tan diversos como pueden ser la satisfacción de los alumnos con la docencia impartida, la atención recibida en los procesos administrativos o el funcionamiento de las aplicaciones institucionales, entre otros. En particular, el personal docente e investigador (PDI) suele recurrir al uso de diversos instrumentos de evaluación e indagación (cuestionarios, escalas de opinión o actitudes, así como pruebas de conocimientos tipo test¹), tanto en su labor docente como investigadora.

Existe una gran variedad de sitios web que ofrecen funcionalidades para la creación y aplicación de encuestas y cuestionarios. Dichos sitios suelen proporcionar facilidades para la gestión y el diseño de formularios *online*, la recogida de respuestas y la generación de informes de resultados. Algunos de ellos además incorporan corrección automática de los test o permiten la generación de gráficos. El uso de estas aplicaciones permite agilizar ciertos procesos y rutinas relacionadas con estas actividades de evaluación e indagación, ahorrando tiempo y esfuerzo.

Sin embargo, el análisis que realizan estas herramientas es, en la mayoría de los casos, limitado. En dicho análisis suele ser habitual el cálculo de estadísticos descriptivos de uso común (frecuencias, porcentajes).

¹ A lo largo del artículo al usar el término *test* haremos referencia a pruebas objetivas de conocimiento basadas en preguntas de respuesta múltiple, también conocidas como pruebas cognitivas.

tajes, medias, etc.). Sin embargo, existen otras métricas o propiedades de los instrumentos de evaluación diseñados que permiten un análisis más profundo (análisis de fiabilidad o del poder discriminativo de los ítems).

Dentro del ámbito de las ciencias sociales existen varias disciplinas que abordan estas cuestiones, entre las que destacamos la Psicometría [1]. Esta área de conocimiento se ocupa de aspectos teóricos y prácticos relacionados con la elaboración, evaluación y aplicación de instrumentos de medición y evaluación. Aunque se encuadra en el área de la medición psicológica, los procedimientos matemáticos que propone, así como los principios de diseño y construcción de escalas de medias y análisis de los test, pueden ser aplicados a una gran variedad de ámbitos.

Así pues, con el objetivo de cubrir algunas de las limitaciones y necesidades detectadas al analizar las herramientas existentes, se ha creado una aplicación web para la gestión, diseño y despliegue de cuestionarios y pruebas de conocimientos tipo test. Uno de los principales aspectos que diferencia la aplicación desarrollada de otras ya existentes es el módulo de análisis de resultados que incorpora, en el que se analizan y calculan gran cantidad de propiedades métricas de los instrumentos de evaluación creados.

La estructura del artículo es la siguiente. En la sección 2 se hace una revisión y comparativa de algunas de las aplicaciones existentes para el diseño y aplicación de encuestas *online*. En la sección 3 se describen las principales propiedades métricas de los instrumentos de evaluación e indagación. La siguiente sección describe las principales funcionalidades soportadas por la aplicación web desarrollada. Finalmente, se comentan las conclusiones extraídas de este trabajo y las líneas de continuación del mismo.

2. Aplicaciones para el diseño y aplicación de encuestas

En la actualidad existen múltiples herramientas, libres o de pago, que permiten generar encuestas y exportar sus resultados. Son aplicaciones muy útiles y demandadas, tanto por empresas como por instituciones, con el fin de obtener información directamente de clientes, estudiantes, trabajadores o cualquier persona implicada en procesos académicos, de formación, de ventas, de captación de clientes, etc. Todos los datos recogidos se estudian posteriormente para sacar conclusiones y definir posibles actuaciones destinadas a mejorar procesos, incentivar compras, resolver deficiencias, etc.

Existe una gran cantidad de sitios web dedicados al diseño de encuestas y recogida de respuestas². Cada

sitio web tienen sus peculiaridades y puede ofrecer alguna característica diferente del resto, pero en general todos ofrecen unos servicios muy similares. En el Cuadro 1 se muestra una comparativa de las principales características soportadas por el siguiente conjunto de herramientas:

- (1) Survey Monkey (<http://es.surveymonkey.com>).
- (2) Survio (<http://www.survio.com/es/>).
- (3) Encuesta Fácil (<http://www.encuestafacil.com/>).
- (4) Online Encuesta (<https://www.onlineencuesta.com/>).
- (5) E-Encuesta (<http://www.e-encuesta.com/>).
- (6) Encuesta Tick (<http://www.portaldeencuestas.com/>).
- (7) Cuestionarios de Moodle³.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
APARIENCIA							
Plantillas	√	√	√	√	√	√	
Logotipo	√	√	√	√	√	√	
Accesibilidad	√						
ENCUESTA							
Indicador progreso	√	√	√		√		
Notificaciones	√	√		√	√	√	√
Exportación resultados	√	√	√	√	√	√	√
Pausar y reanudar			√	√	√		√
Tiempo limitado		√		√	√	√	√
Contenido multimedia		√	√	√	√		√
PREGUNTAS							
Diversos tipos de preguntas	√	√	√	√	√	√	√
Validación de respuestas	√		√				√
RESULTADOS							
Exportar resultados	√	√	√	√	√	√	√
Informes con gráficos	√	√	√	√	√	√	
Estadísticos básicos							√
Cálculo de Fiabilidad							
Propiedades métricas de los test							√

Cuadro 1: Comparativa de sitios web para la gestión y despliegue de encuestas.

La aplicación creada y descrita en este artículo incorpora todas las funcionalidades incluidas en la

² Existe gran número de aplicaciones de uso más genérico que también incorporan la creación de formularios (como las de

Google) y que pueden usarse para la creación, despliegue y recogida de resultados de cuestionarios.

³ <https://docs.moodle.org/all/es/19/Cuestionarios>

tabla comparativa. Entre las características que la diferencian del resto, destacamos la posibilidad de calcular, además de los estadísticos descriptivos habituales, el cálculo de las métricas de calidad de los instrumentos de medición creados.

3. Propiedades métricas de los instrumentos de evaluación e indagación docente

Según la Real Academia Española, se puede definir Psicometría como la “*medida de los fenómenos psíquicos*”. Se trata de una disciplina metodológica encargada de la medición de variables psicológicas de interés, para lo que proporciona métodos para evaluar diferentes propiedades de los instrumentos de medición, como los *cuestionarios* o *escalas* (para la medición de actitudes), y los *test* (para la evaluación de conocimientos y competencias) [1].

Los cuestionarios y los test, como otros instrumentos de medida, no nos proporcionan una medida perfecta, por lo que existe un elemento de error e imperfección que resulta imprescindible conocer para saber si mide lo que pretende medir y la precisión con que lo hace. Por eso, uno de los aspectos más importantes en cualquier evaluación es saber elegir los instrumentos utilizados según criterios de calidad [6]. Los principales criterios de calidad o bondad asumidos como normas en la construcción, interpretación y utilización de instrumentos de medición son la *fiabilidad* y la *validez* [8]. La Psicometría desarrolla y estudia procedimientos estadísticos destinados a comprobar si una prueba es válida y fiable para medir una determinada variable [1].

Algunas de las propiedades métricas de los instrumentos de evaluación e indagación que son de interés en tareas de evaluación e investigación docente, son las siguientes:

- **Fiabilidad del test⁴**. Es una propiedad métrica que hace referencia al grado de consistencia y estabilidad de las puntuaciones obtenidas por los sujetos a lo largo de sucesivas pruebas de medición mediante un mismo instrumento de medida. Por tanto, indica en qué medida, en pruebas similares, los encuestados hubieran quedado ordenados de manera parecida.

Un coeficiente de fiabilidad alto indica que el test ha logrado ordenar la muestra de sujetos de acuerdo a sus respuestas. Por ejemplo, un coeficiente de fiabilidad alto en un examen tipo test indica que se ha logrado ordenar la muestra de

sujetos de acuerdo a sus conocimientos. La fiabilidad de los test se mide mediante el coeficiente *Alpha* de Cronbach (α) [3]. Este coeficiente se puede calcular tanto para pruebas cognitivas o test de rendimiento con ítems dicotómicos (acierto/error), como para pruebas no cognitivas con ítems no dicotómicos (escalas de Likert) [5]. Su valor oscila entre 0 y 1, siendo considerados los índices superiores a 0,7 suficientes para garantizar la fiabilidad de un test.

- **Análisis de los ítems individuales de un test de conocimientos**. Dicho análisis se basa en el cálculo de los siguientes índices: (a) el *índice de dificultad del ítem* (ID), que hace alusión a la proporción de sujetos que lo responden correctamente y (b) el *índice de dificultad del ítem corregido* (IDc), que tiene en cuenta la suerte de aquellos sujetos que sin conocer la respuesta eligen la alternativa correcta.

El índice de dificultad de un ítem (ID) puede tomar valores entre 0 y 1. Un índice próximo a 0 indica que el ítem tiene una dificultad baja, mientras que cuanto más se acerque a 1, mayor será su dificultad. Aquellos ítems con valores extremos deben retirarse del test, ya que casi todos los alumnos lo aciertan o casi todos lo fallan, y no sirven para diferenciar a los sujetos con diferentes niveles del rasgo medido, es decir, con diferentes niveles de competencia.

Partiendo de la idea de que los sujetos con buenas puntuaciones en un test deben acertar un determinado ítem en mayor proporción que los sujetos con bajas puntuaciones, surge la necesidad del cálculo del *poder discriminativo de un ítem*. Esto es, un ítem debe poder discriminar entre sujetos con distintos niveles de competencia.

- **Poder discriminativo de un ítem**. indica la capacidad para diferenciar a los sujetos con distinto nivel en el rasgo medido. El poder discriminativo de un ítem se puede estudiar de distintas formas: (a) como la proporción de sujetos que acierta el ítem (PROP), y que coincide con el índice de dificultad sin corregir y (b) el cálculo de la correlación biserial-puntual del ítem-test (RPBI).

La correlación biserial-puntual de un ítem indica en qué medida el hecho de responder correctamente una pregunta está relacionado con obtener una buena puntuación en el test. Una correlación cercana a cero indica que el responder bien o mal a la pregunta no es indicador de obtener una buena o mala puntuación en el test. Una correlación negativa indica que el responder bien a la pregunta es indicador de obtener una mala puntuación en el test (por tanto, se trata de una pregunta mal diseñada). Por último, una correlación positiva indica que el responder bien a la pregunta sí

⁴ En muchas ocasiones se confunde la *fiabilidad* con la *validez*. Un test es fiable si permite ordenar a los sujetos según sus respuestas o nivel de conocimientos (en el caso de las pruebas cognitivas). Un test es válido si permite medir aquello para lo que ha sido diseñado.

es un indicador de obtener una buena puntuación en el test (se trata de una buena pregunta). Por lo tanto, las preguntas con RPBI positivo son las que mejor discriminan la muestra de sujetos.

- **Análisis de los distractores.** Tan importante como calcular el poder discriminativo de un ítem, lo es el calcular el poder discriminativo de cada distractor. Se conoce como *distractor* a cada una de las opciones de respuesta incorrectas de un ítem aptitudinal o de rendimiento (incluido en un test de conocimientos o prueba cognitiva). El análisis de los distractores permite conocer la utilidad de cada alternativa incorrecta en una cuestión de respuesta múltiple y, por tanto, su contribución a la calidad del ítem. También es útil para averiguar si alguna alternativa no ha sido seleccionada por ningún sujeto, o una proporción muy baja de ellos.

En dicho análisis se debe tener en cuenta que cada distractor debe ser elegido por un mínimo de sujetos (al menos por un 10% de la muestra). Además es recomendable que este porcentaje sea lo más parecido posible entre los diferentes distractores. Además, la puntuación obtenida por los sujetos que seleccionan la opción incorrecta debería ser menor a la obtenida por quienes aciertan la pregunta. Por último, a medida que el nivel de aptitud de los sujetos aumente, se espera que disminuya el porcentaje de selección de los distractores y viceversa.

El poder discriminativo de un distractor se puede estudiar de distintas formas: (a) calculando la proporción de sujetos que selecciona el distractor (PROP) y la correlación biserial-puntual del distractor-test (RPBI).

El método de cálculo es similar al utilizado en el análisis de los ítems individuales de los test. Lo esperado es que el RPBI sea positivo para la opción correcta y negativo para los distractores.

El análisis de los ítems individuales y los distractores proporciona una serie de ventajas, entre las que destacamos:

- *Mejora la calidad del test:* aportan información sobre cada una de las preguntas, permitiendo descartar aquellas que no discriminan correctamente la muestra de sujetos. Es una forma de autoevaluación del profesor y del diseño del test.
- *Aporta información útil:* se pueden extraer conclusiones sobre el test realizado y ofrecer *feedback* sobre los aciertos y errores más frecuentes.
- *Detecta posibles errores en la formulación:* como pueden ser preguntas ambiguas, varias respuestas correctas, clave de corrección equivocada, nivel de dificultad excesivo o insuficiente, etc.

- *Fácilmente programables:* son procedimientos sencillos de calcular, que aportan información muy valiosa con un coste reducido en tiempo y esfuerzo.

El uso y posterior análisis de los ítems a lo largo de varios cursos permitirá a los docentes crear una base de datos o batería de preguntas tipo test, clasificadas según su nivel de dificultad y poder discriminatorio, que pueden ser reutilizadas en distintas convocatorias y cursos posteriores. Conocer la calidad métrica de dichas preguntas o ítems permite crear pruebas tipo test mejor diseñadas y con una proporción mejor balanceada de preguntas con distinto grado de dificultad.

4. La herramienta *QualitySurvey*

Tal y como se ha comentado anteriormente nuestro objetivo era la implementación de una aplicación web que permitiera: (a) el diseño de cuestionarios y pruebas cognitivas o de evaluación de competencias tipo test, (b) la recopilación de respuestas por parte de los alumnos o encuestados y (c) que incorporase un completo módulo de análisis de los resultados obtenidos, en el que se incluyera el cálculo de las métricas de calidad de los instrumentos de evaluación e indagación diseñados.

La herramienta *QualitySurvey*⁵ soporta funcionalidades destinadas a dos roles: el gestor y el encuestado. El *gestor* será el encargado de diseñar encuestas, enviar las invitaciones a los encuestados y generar los informes de resultados (Figura 1). Por su parte, el *encuestado* puede consultar las encuestas que tiene pendiente de contestar, así como cumplimentar encuestas o test, previa invitación.

La aplicación desarrollada da soporte a tres tipos de instrumentos de evaluación e indagación:

- *Escalas de opinión o actitudes* (pruebas no cognitivas). Este instrumento de recogida de datos consta exclusivamente de preguntas de tipo *escala de Likert* [5], que podrán ser relativas a distintos aspectos relacionados con la docencia y/o la investigación. Podrán emplearse para conocer la percepción subjetiva de los encuestados en relación a determinados aspectos relacionados con la calidad de la docencia, con aspectos concretos relacionados con los contenidos impartidos o la metodología empleada en clase, entre otros.
- *Cuestionarios* (pruebas no cognitivas). Con un objetivo similar al tipo anterior, incorporará una mayor variedad de tipos de preguntas (sí/no, lista de elementos, texto,...), que permitan un análisis más completo de la opinión de los encuestados.

⁵ <http://www.uclm.es/varios/encuestas/adm/>

- **Test de conocimiento** (pruebas cognitivas). Pruebas de conocimientos compuestas por un conjunto de preguntas de respuesta múltiple, en las que una de las opciones será la correcta y el resto distractores. Se podrán usar para la evaluación objetiva de competencias.

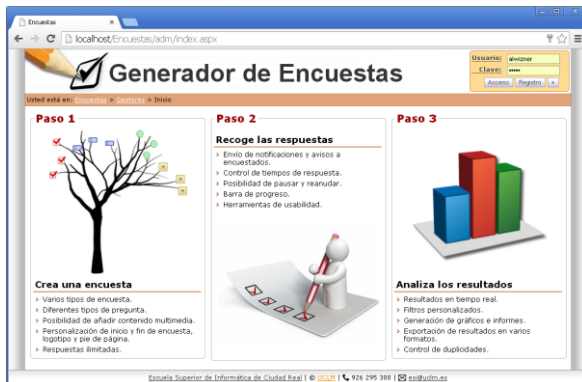


Figura 1: Aspecto de la página de inicio con la que interactúan los gestores.

Haremos uso del término genérico “*encuesta*” para hacer referencia a los tres tipos de instrumentos de medición soportados. Las encuestas creadas estarán asociadas a *convocatorias* o periodos en los que las encuestas están disponibles (Figura 2.a). Para diseñar la encuesta se deben crear *módulos* (o conjuntos de *preguntas*) (Figura 2.b). Las preguntas podrán ser de distinto tipo (de dos alternativas, tipo escala, de elección múltiple, etc.) (Figura 2.c), siendo posible asociarles *contenidos multimedia* (videos, imágenes, audios, etc.) (Figura 2.d).

La aplicación permite tanto diseñar encuestas desde cero como crear una nueva encuesta a partir de una *plantillas* prediseñadas. El módulo de creación de encuestas además incluye enlaces a recomendaciones o guías relativas a la confección de los ítems (redacción de los mismos, formato de las preguntas, confección de los distractores y la respuesta correcta, etc.), así como recomendaciones relativas a la descripción de los recursos multimedia asociados a las preguntas (Figura 2.d).

En relación a las encuestas no solo se puede especificar el periodo en el que estarán activas, sino también la posibilidad de fijar un tiempo máximo para su cumplimentación.

En cuanto a la *apariciencia* de la encuesta, la aplicación permite la personalización de logotipo y pie de página, la especificación del número de módulos que se mostrarán por cada página, así como la posibilidad de añadir una página inicial y final a la encuesta. El contenido de estas páginas puede crearse o modificarse a partir de un repositorio de textos predefinidos disponibles en la herramienta (instrucciones para cumplimentar la encuesta, la nota de confidencialidad o un mensaje de agradecimiento, entre otros).



Figura 2: (a) Aspecto de la interfaz de usuario para la gestión de *convocatorias* (b) de *encuestas* y (c) *preguntas* o ítems individuales (d) Ventana modal de inserción de *recursos multimedia* (incluyendo recomendaciones sobre la descripción asociada).

En la Figura 3 se puede ver el aspecto de una encuesta creada y disponible para ser cumplimentada por el encuestado. La aplicación soporta aspectos de *accesibilidad* web (incorporando la posibilidad de

cambiar el contraste, el tamaño de la fuente, etc.) [2], disponibles en la barra de herramientas que se muestra en la parte superior de la encuesta. En dicha barra también se puede ver el tiempo restante para cumplimentar la encuesta (en aquellas en las que se haya fijado un tiempo máximo para su cumplimentación), así como el botón que permite pausar/reanudar la encuesta. Por último, se puede ver la barra de progreso que indica el porcentaje de encuesta completado.

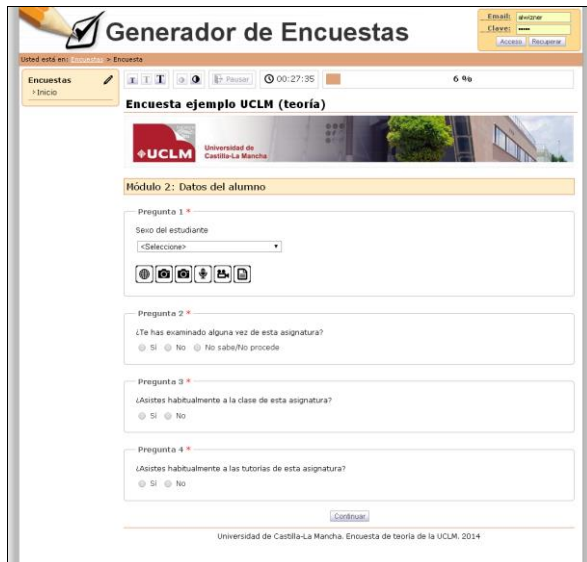


Figura 3: Aspecto final de la encuesta generada.

El sistema incorpora un completo sistema de gestión y seguimiento de *notificaciones* a encuestados (invitaciones y recordatorios, plantillas para su creación, etc.), así como la posibilidad de consultar el histórico de encuestas y notificaciones.

Una vez que los encuestados hayan completado sus encuestas (questionarios o test), la aplicación permite la exportación de sus respuestas en distintos formatos (HTML, PDF, DOC, XLS y CSV). En el caso de los test de conocimientos, también se habilita una opción adicional para la *descarga de las calificaciones obtenidas* en estos mismos formatos (Figura 4).

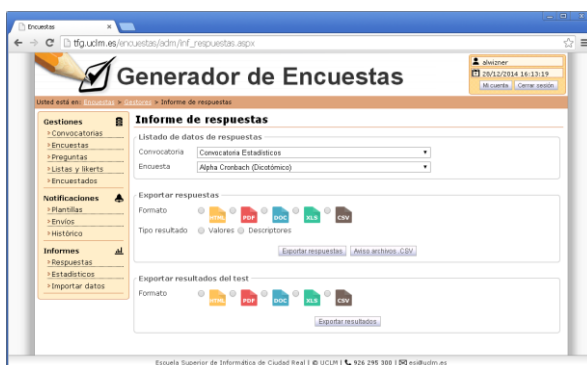


Figura 4: Funcionalidad de descarga de respuestas y resultados.

Tal y como se ha indicado, uno de los aspectos que diferencia la aplicación desarrollada del resto de aplicaciones similares existentes en el mercado es el módulo de análisis de resultados. Dicho módulo incluye, además del cálculo de los estadísticos descriptivos habituales, el análisis detallado de las propiedades métricas de los instrumentos de indagación y evaluación diseñados.

Dependiendo del tipo de encuesta creada (de satisfacción, de opinión o prueba de conocimiento) se calcularán distintos estadísticos y métricas. La aplicación permite la creación y exportación de informes y gráficas de resultados en los formatos antes comentados. En primer lugar, el informe muestra el valor del *Alpha* de Cronbach (Figura 5), indicativo de la fiabilidad de la encuesta.

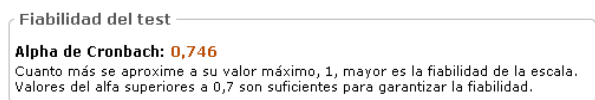


Figura 5: Ejemplo de *Alpha* de Cronbach en el informe generado para una encuesta de opinión.

En la Figura 6 podemos ver un extracto del informe generado a partir de los datos recogidos mediante una *escala de opinión* (formada exclusivamente por escalas de Likert). En este caso se calculan los estadísticos habituales (media, mediana, moda y desviación típica) y, además, se generan gráficos de barras y de tipo radar para cada una de las preguntas que componen la encuesta.

Al final del informe también se incorporan otros dos gráficos de barras y de tipo radar con los resultados totales por cada módulo en los que se divide la encuesta de opinión (Figura 7).

En los informes generados en el caso de los *test de conocimientos* (Figura 8) se muestra, para cada ítem (pregunta tipo test) el índice de dificultad (ID) del ítem, así como el índice de dificultad corregido (IDc). Además, en función de los valores de estos índices, las preguntas son clasificadas (utilizando un código de colores) en fáciles, de dificultad media y difíciles. Por último se incluyen las métricas relativas al poder discriminativo del ítem y de sus distractores (PROP y RPBI), así como un conjunto de iconos que resaltan situaciones anómalas, como distractores seleccionados por menos del 10% de alumnos y valores del RPBI de signo contrario al esperado (Cuadro 2).

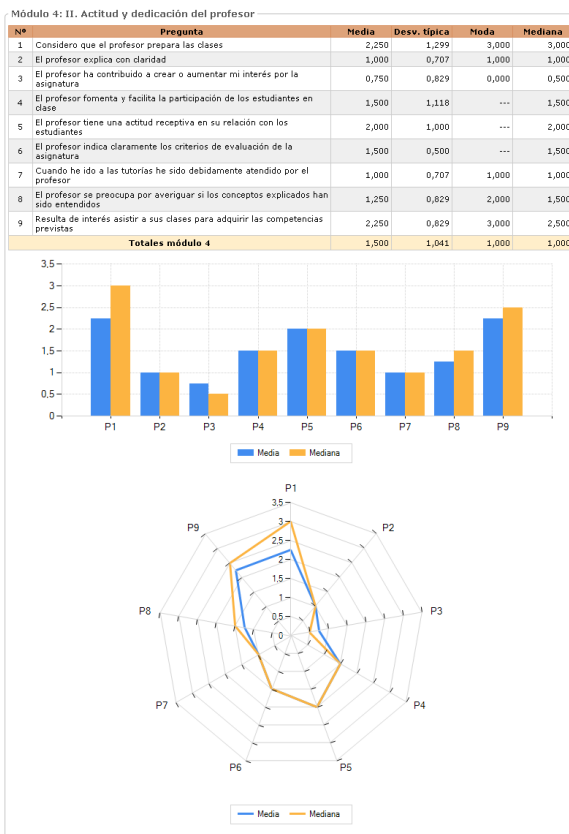


Figura 6: Detalle del informe de un módulo de una encuesta de opinión (incluye estadísticos y gráficas).

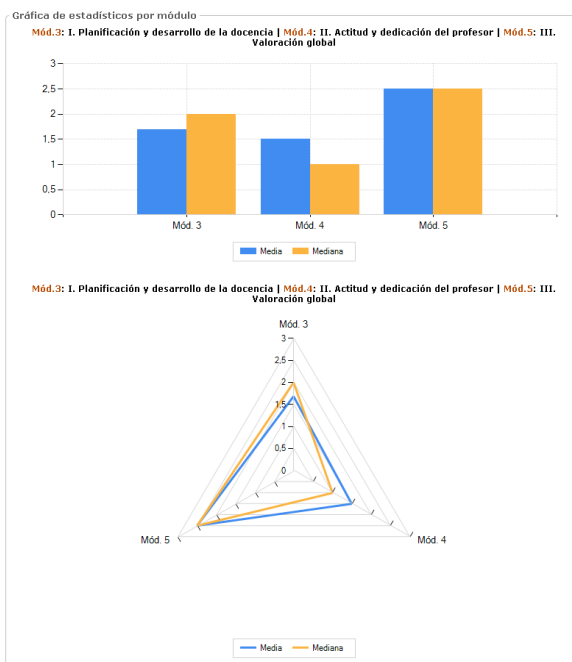


Figura 7: Detalle de los gráficos generados para el conjunto de módulos en que se divide una encuesta de opinión.

Icono	Índice	Representación
F	ID	Pregunta considerada de dificultad baja.
M	ID	Pregunta considerada de dificultad media.
D	ID	Pregunta considerada de dificultad alta.
⚠	PROP	Aparece en las opciones de una pregunta que se pueden considerar como un mal distractor, debido a que no han sido seleccionadas por, como mínimo, un 10% de los encuestados.
	RPBI	Aparece en los distractores de una pregunta cuyo RPBI es positivo. Aparece en la respuesta correcta de una pregunta cuyo RPBI es negativo.
Σ	PROP	Aparece en la respuesta correcta de una pregunta considerada de alta dificultad, debido a que el porcentaje de encuestados que han seleccionado cualquiera de los distractores es superior que el porcentaje de encuestados que han respondido la opción correcta.

Cuadro 2: Interpretación de iconos mostrados en los informes asociados a los test de conocimientos.

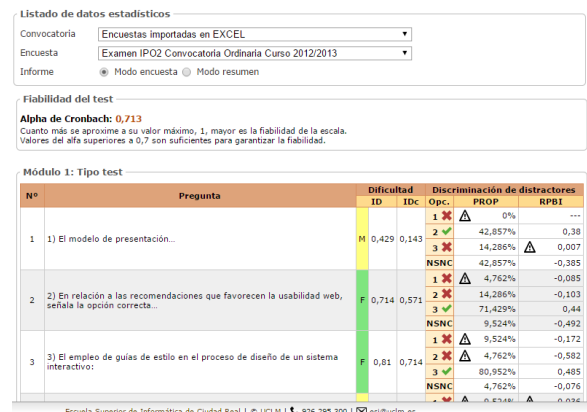


Figura 8: Detalle del informe generado para una prueba de conocimientos tipo test⁶.

Los informes generados pueden ser completos o en *modo resumen* (Figura 9). En este caso aparecen agrupadas las preguntas por nivel de dificultad, pudiendo estas ser fáciles (D), de dificultad media (M) o fáciles (F). También se incluye el listado de aquellos ítems o preguntas cuya métrica de calidad presenta un valor anómalo.

Por último, la aplicación también incluye la funcionalidad de importar respuestas de encuestas y test en formato Excel. Esta última funcionalidad será de interés para profesores que quieran usar la herramienta sólo para comprobar la calidad de sus test y encuestas, pero que hayan sido aplicados sin hacer uso de la herramienta desarrollada (por ejemplo, en el caso de pruebas en papel o que se

⁶ Extraído a partir de datos reales obtenidos en la evaluación de la asignatura Interacción Persona-Ordenador II del Grado de Ingeniería Informática de la UCLM-Ciudad Real (curso académico 2012/2013).

hayan llevado a cabo antes del desarrollo de esta aplicación). Para ello la aplicación permite descargar la plantilla Excel a cumplimentar (con un formato específico) para que los datos sean interpretables por la aplicación. Una vez hecha la importación podrá realizarse el análisis de resultados y de las propiedades del test, pudiendo generarse los informes y gráficas correspondientes.

Preguntas con dificultad alta		
Hód.	Preg.	Pregunta
1	3	Pregunta 3
1	5	Pregunta 5
1	7	Pregunta 7
1	12	Pregunta 12
1	14	Pregunta 14
1	16	Pregunta 16

Preguntas con dificultad media		
Hód.	Preg.	Pregunta
1	1	Pregunta 1
1	2	Pregunta 2
1	4	Pregunta 4
1	6	Pregunta 6
1	11	Pregunta 11
1	13	Pregunta 13
1	15	Pregunta 15

Preguntas con dificultad baja		
Hód.	Preg.	Pregunta
1	8	Pregunta 8
1	9	Pregunta 9
1	10	Pregunta 10

Preguntas con % selección distractores > % opción correcta				
Hód.	Preg.	Pregunta	Aciertos	Distractores
1	3	Pregunta 3	18,367%	51,021%
1	5	Pregunta 5	18,367%	42,857%
1	12	Pregunta 12	16,327%	30,612%
1	14	Pregunta 14	32,653%	51,021%
1	16	Pregunta 16	18,367%	75,510%

Preguntas con distractores contestados por < 10%				
Hód.	Preg.	Pregunta	Opc.	PROP.
1	1	Pregunta 1	3	2,041%
1	4	Pregunta 4	1	8,163%
1	6	Pregunta 6	3	2,041%
1	7	Pregunta 7	3	8,163%
1	8	Pregunta 8	2	0%
1	9	Pregunta 9	3	0%
1	9	Pregunta 9	3	6,122%

Figura 9: Detalle del informe generado en modo resumen⁷.

5. Conclusiones

En este artículo se ha descrito una herramienta para el diseño, gestión y despliegue de encuestas y test de conocimientos. La aplicación también permite la gestión integral de usuarios, gestión de invitaciones a encuestas, análisis de resultados, generación de informes, descarga de resultados en varios formatos y generación de gráficos, entre otros. En su diseño e implementación se han tenido en cuenta aspectos de usabilidad y accesibilidad web [4, 7].

Uno de los aspectos más destacados de la aplicación desarrollada es el módulo de análisis de cuestionarios e ítems. Su uso proporciona una serie de ventajas entre las que destacan:

- Posibilidad de testear la calidad de los instrumentos de medición y evaluación diseñados.
- Detección de errores de formulación de ítems.

- Posibilidad de creación de un repositorio de cuestionarios e ítems, que podrán aplicarse en distintas convocatorias.
- Creación de baterías de preguntas tipo test, clasificadas por su nivel de dificultad y poder discriminativo, que permitirán crear test de evaluación de conocimientos mejor diseñados y balanceados en el futuro, etc.

Además, la posibilidad de importar resultados externos permitirá el uso de la aplicación para analizar pruebas o cuestionarios realizados con anterioridad al desarrollo de la misma.

Por lo tanto creemos que el recurso desarrollado puede ser de gran utilidad en el ámbito educativo. Los usuarios finales de dicha aplicación podrán ser alumnos, profesores y miembros de comisiones de evaluación de la calidad docente y/o investigadora, que podrán emplearla en actividades de seguimiento, así como de evaluación de competencias.

Entre las propuestas de trabajo futuro nos planteamos incorporar el soporte a nuevos tipos de preguntas (cloze, emparejamiento, etc.), la importación y exportación de preguntas (en un formato de preguntas test, como GIFT), así como la integración del recurso desarrollado en algún entorno de aprendizaje de uso extendido (que ya incorpore la gestión de usuarios), como Moodle⁸.

Referencias

- [1] M. Barbero, E. Vila, y F. Holgado. Psicometría. Editorial Sanz y Torres, 2010.
- [2] W. Chisholm y M. May, Universal Design for Web Applications: Web Applications That Reach Everyone, 2008.
- [3] L. Cronbach, Coefficient alpha and internal structure of test, *Psychometrika*, vol. 16, pp. 297-334, 1951.
- [4] C. Jarrett y G. Gaffney, Forms that Work. Designing Web Forms for Usability. Editorial Morgan Kaufman Publ Inc, 2008.
- [5] R. Likert, A technique for the measurement of attitudes, *Archives of Psychology*, vol. 22, nº 140, pp. 44-53, 1932.
- [6] I. Lucero y S. Meza, Validación de instrumentos para medir conocimientos, Argentina, 2002.
- [7] J. Nielsen, Prioritizing Web Usability, Editorial New Riders, 2006.
- [8] G. D. A. Prieto, Fiabilidad y Validez, *Papeles del Psicólogo*, vol. 31, nº 1, pp. 67-74, 2010.

⁷ Extraído a partir de datos reales obtenidos en la evaluación de la asignatura Interacción Persona-Ordenador II del Grado de Ingeniería Informática de la UCLM-Ciudad Real (curso académico 2013/2014).

⁸ <https://moodle.org>