



Análisis de la fiabilidad y validez de un cuestionario docente

Carmen Lacave Rodero¹, Ana Isabel Molina Díaz¹, Mercedes Fernández Guerrero²,
Miguel Ángel Redondo Duque¹

¹Dpto. de Tecnologías y Sistemas de la Información

²Departamento de Matemáticas

Universidad de Castilla La Mancha

carmen.lacave@uclm.es, anaisabel.molina@uclm.es, mercedes.fernandez@uclm.es, miguel.redondo@uclm.es

Resumen

En la actualidad los cuestionarios constituyen el instrumento más utilizado por el profesorado para evaluar distintos aspectos relacionados con la docencia ya que el análisis estadístico de los datos recogidos a través de los mismos permite inferir conclusiones a los docentes de forma sencilla y rigurosa. Sin embargo, en este proceso se suele pasar por alto el hecho de que, para garantizar la utilidad y significado de los resultados obtenidos, no es suficiente la simple presentación de una lista de preguntas diseñada *ad hoc*, sino que el cuestionario utilizado debe estar bien diseñado. Los principales criterios de calidad exigibles a los instrumentos de evaluación son la fiabilidad y la validez. En este trabajo describimos el análisis de la validez y fiabilidad de un cuestionario para medir las dificultades de los estudiantes durante el aprendizaje de la recursividad. Concluimos que este estudio permite proporcionar a la comunidad educativa una herramienta de evaluación de rápida aplicación para valorar el problema de la recursividad. Además, el proceso de análisis descrito a lo largo del artículo puede extrapolarse fácilmente, por lo que consideramos que puede servir como guía para la validación de cualquier otro cuestionario docente.

Palabras clave: Cuestionario, calidad, fiabilidad, validez, recursividad.

Recibido: 17 de noviembre de 2015; **Aceptado:** 21 de diciembre de 2015.

1. Introducción

Antes de la implantación del Espacio Europeo de Educación Superior (EEES) [7] la docencia universitaria ha estado basada mayoritariamente en la impartición de clases magistrales, mediante las que el profesor realiza la transmisión de sus conocimientos a sus alumnos, quienes toman apuntes y asimilan los conceptos. Sin embargo, los principios que promueve el EEES [4] exigen al profesor rediseñar el modelo docente tradicional, haciendo uso de propuestas innovadoras en aras de la calidad docente y, en consecuencia, del proceso y los resultados de aprendizaje de los alumnos [26, 27].

Es indudable que, en este contexto, todo cambio debe incorporar un proceso de evaluación de su efectividad, con el fin de decidir si procede mantenerlo o no [37]. El método más sencillo y rápido que se suele utilizar para evaluar distintos aspectos relacionados con la docencia es el uso de cuestiona-

rios¹ [24].

Como ilustra la figura 1, cuando el docente decide medir el efecto que produce la aplicación de una innovación educativa en el aprendizaje de sus estudiantes, normalmente elabora uno o varios cuestionarios que posteriormente contestan los alumnos implicados en la experiencia. Los datos proporcionados por estos cuestionarios son analizados mediante técnicas estadísticas básicas y se utilizan para difundir los resultados en conferencias o publicaciones especializadas con el fin de que, si los resultados son prometedores, la experiencia se pueda replicar en otro contexto.

Pero para garantizar la utilidad y significado de los resultados obtenidos, no vale con la simple creación de una lista de preguntas y respuestas, sino que el instrumento utilizado debe estar bien “calibrado”, es decir, debe estar bien diseñado según los criterios estándares de calidad. Los principales criterios de calidad o bondad asumidos como normas en la construcción,

¹A lo largo del trabajo cuando hablamos de *cuestionario* nos referimos a un instrumento de medición formado por una colección de preguntas con respuestas cerradas y ordenadas según alguna escala.



Figura 1: Proceso de evaluación de una innovación docente.

interpretación y utilización de instrumentos de medición son la *validez* y la *fiabilidad* [30]. La validez se refiere al grado en que el instrumento mide lo que se pretende medir; la fiabilidad de un cuestionario se refiere a la confianza que se concede a los datos que se obtienen con el mismo y está relacionada con la coherencia o consistencia interna y la precisión de las medidas recopiladas. Estas dos condiciones son cruciales, porque si el cuestionario es “defectuoso”, es decir, no tiene la calidad mínima exigible, no se puede garantizar el éxito del diagnóstico, ya que ningún tratamiento estadístico sobre datos de mala calidad puede generar y garantizar buenos resultados [23]. Merece la pena destacar que la fiabilidad y la validez no son características de los cuestionarios [30], sino que corresponden a propiedades de las interpretaciones, inferencias o usos específicos de las medidas que proporcionan los cuestionarios, debiendo entenderlas además como una cuestión de grado [1]. Esto implica que la validez y la fiabilidad de un cuestionario no son cualidades que se puedan definir en términos absolutos sino que hay que entenderlas en relación con los valores que han arrojado los métodos de análisis de la validez y fiabilidad del instrumento. En consecuencia, un mismo instrumento puede medir o clasificar bien a los sujetos de una muestra, con mucha precisión, y mal, con un margen de error grande, a los sujetos de otra muestra. Por tanto, la validez y la fiabilidad se deben calcular con cada nueva muestra, sin aducir los resultados obtenidos con otras muestras como aval de la calidad del instrumento [28].

Aunque existen multitud de cuestionarios aceptados por la comunidad científica en distintos ámbitos como la medicina, psicología, etc., la educación no se ha articulado como un campo basado en evidencias, en parte debido al cuestionable impacto de la investigación sobre la práctica docente [8]. No obstante, la evaluación de la calidad docente es un campo, cada vez más en auge, en el que se viene trabajando desde el último cuarto del siglo XX, por lo que existen algunos cuestionarios muy utilizados a nivel internacional, como el SEEQ [25] o el ENDEAVOR [14], que permiten analizar la eficacia de la enseñanza utilizando una serie de factores. La elección de estos cuestionarios está fundamentada en tres ventajas [35]:

- sus propiedades psicométricas. La psicometría es la disciplina que desarrolla y estudia procedimientos estadís-

ticos destinados a comprobar si una prueba es válida y fiable para medir una determinada variable [16];

- su amplia utilización en universidades de todo el mundo, tanto en su forma original como adaptados de algún modo [33]; y
- la gran cantidad de material disponible para el perfeccionamiento de cada uno de los ítems analizados [35].

En otras ocasiones en las que es necesaria una evaluación de aspectos más concretos, el profesorado se ve en la obligación de diseñar su propio instrumento de medida. Esta tarea no es sencilla pues construir un cuestionario técnicamente bien hecho conlleva, en sí mismo, una investigación [31]. Por tanto, el uso de cuestionarios docentes implica necesariamente el control de su calidad mediante el estudio de su validez y fiabilidad. Sin embargo, los profesores normalmente carecemos de los conocimientos estadísticos (o psicométricos) necesarios para realizar tanto el diseño como la validación de la calidad de los cuestionarios diseñados por nosotros mismos. Este hecho ha motivado que nos planteásemos describir las bases conceptuales de la psicometría para la validación de un cuestionario docente, mediante su aplicación al cuestionario CoMaR (Compresión y Manejo de la Recursividad), diseñado por nosotros mismos, para la identificación de las dificultades de los alumnos durante el proceso de aprendizaje de la recursividad [20].

La recursividad es una técnica de programación muy potente, puesto que permite resolver problemas muy complejos en función de las soluciones de los mismos problemas, pero de menor tamaño, basándose en el principio de inducción matemática. Es un concepto básico, que aparece en todos los currículos de Informática, en los primeros cursos de programación. Sin embargo, constituye uno de los conceptos más difíciles de entender por los alumnos que están aprendiendo a programar. Tanto es así, que desde hace tiempo se trabaja intentando conocer las principales causas por las que resulta tan complicado dominarla [32, 34] pero no existen trabajos que hayan ofrecido datos empíricos que justifiquen el porqué de utilizar unas u otras herramientas de visualización o que justifiquen el empleo de una metodología específica. Además,

las propuestas que se han realizado no inciden en las distintas necesidades del alumnado. Por tanto, nos planteamos diseñar un cuestionario mediante el que pudiésemos identificar los elementos o características de la recursividad en los que es necesario incidir más dependiendo de las necesidades de cada estudiante según la etapa del aprendizaje en la que se encuentre [20].

En consecuencia, en este trabajo describimos el proceso de análisis de la calidad de dicho cuestionario, mediante el estudio de su validez y fiabilidad, con dos objetivos:

- definir un método claro y sencillo de calibración de cuestionarios, extrapolable a otros contextos docentes, y
- proporcionar a la comunidad docente una herramienta de rápida aplicación y evaluación, de bajo coste y utilizable en la mayor parte de contextos posibles, que permita valorar la magnitud del problema de la recursividad.

2. Análisis de la calidad de un cuestionario docente

Considerando que el análisis de la calidad de un cuestionario constituye una experiencia de investigación, el estudio debe comenzar con la descripción de sus *características* externas, entre las que destacamos las siguientes:

- *Sujetos* que contestan el cuestionario en cuestión, indicando todos los datos relevantes para el estudio, como su número, edad promedio, desglose por sexos, número de alumnos que repiten curso o asignatura, centro al que pertenecen, titulación y curso que estudian y asignatura, en su caso, en la que se desarrolla la experiencia.
- *Temporalidad*, indicando el momento y duración de la misma.
- *Contenido*, justificando en su caso, la definición de las distintas *dimensiones o factores*² en las que se estructuran o agrupan los ítems o preguntas que componen el cuestionario. Además, se debe indicar el número y tipo de respuestas de cada uno.
- *Herramientas* utilizadas para realizar el análisis del cuestionario.

Además, un cuestionario no nos proporciona una medida perfecta, sino que existe un elemento de error o imperfección asociado al mismo que resulta imprescindible conocer para saber si mide lo que pretende medir, es decir, su validez, y la precisión con que lo hace, esto es, su fiabilidad. En la figura 2 se resumen los aspectos más relevantes de la medición

de ambas características, descritas con detalle en las siguientes secciones. Sin embargo hay que destacar el hecho de que ambas fases se llevan a cabo dentro de un proceso incremental e iterativo en el sentido de que aunque se suele comenzar analizando la validez del cuestionario, el análisis posterior de su fiabilidad puede conllevar la eliminación de alguno de sus ítems, por lo que es recomendable realizar después un análisis de validez para su confirmación posterior.

2.1. Validez

Para el estudio de la validez de un cuestionario docente se recomienda la realización de los siguientes análisis:

Validez de contenido del cuestionario, con el fin de determinar el grado de comprensión de las preguntas de las que consta el cuestionario. A diferencia de otros tipos de validez, la de contenido no suele ser expresada cuantitativamente a través de un índice o coeficiente sino que, por lo general, se estima de manera subjetiva o intersubjetiva. El procedimiento más comúnmente empleado para determinar este tipo de validez, es el que se conoce con el nombre de juicios de expertos, mediante el que se realiza una valoración por personas calificadas en el tema [9]. No obstante, en el caso en el que no exista acuerdo entre los jueces, se puede calcular el índice de validez de contenido [21], que permite cuantificar dicha validez basándose en la valoración de cada uno de los jueces.

Validez de constructo, con el objetivo de explorar el cuestionario y averiguar si las relaciones entre las variables³ definen una estructura dimensional en el cuestionario que se mantenga invariante y pueda servir de base para realizar una interpretación con fundamentos sólidos de los resultados en distintas poblaciones [30]. Se entiende como constructo un concepto que representa una cualidad o atributo de los sujetos, de naturaleza más o menos abstracta, y que se supone representado mediante un modelo matemático en las puntuaciones de un determinado instrumento [33]. El establecimiento de modelos juega un papel primordial para establecer una correspondencia entre el sistema empírico (el mundo real) y un sistema formal (el modelo), ya que el conocimiento preciso de los procesos de la educación sólo es posible con el apoyo en medidas correctas y adecuadas de las variables relevantes [33].

Previamente, y para comprobar la pertinencia de este tipo de análisis, se recomienda calcular la *medida de adecuación muestral* (KMO) [19], que contrasta si las correlaciones parciales entre las variables son suficientemente pequeñas. El estadístico KMO varía entre 0 y 1; un valor menor que 0.5 se interpreta como que la correlación entre dichas variables no es suficientemente significativa, por lo que no tendría sentido realizar un análisis de las relaciones entre las variables con los datos muestrales que se están utilizando [16]. Ade-

²Las dimensiones o factores de un cuestionario representan los aspectos que se desean medir y determinan los puntos sobre los que obtener información mediante los correspondientes ítems del cuestionario.

³El término «variables» es una forma abreviada de referirnos a las variables asociadas a las preguntas o ítems contenidos en el cuestionario.

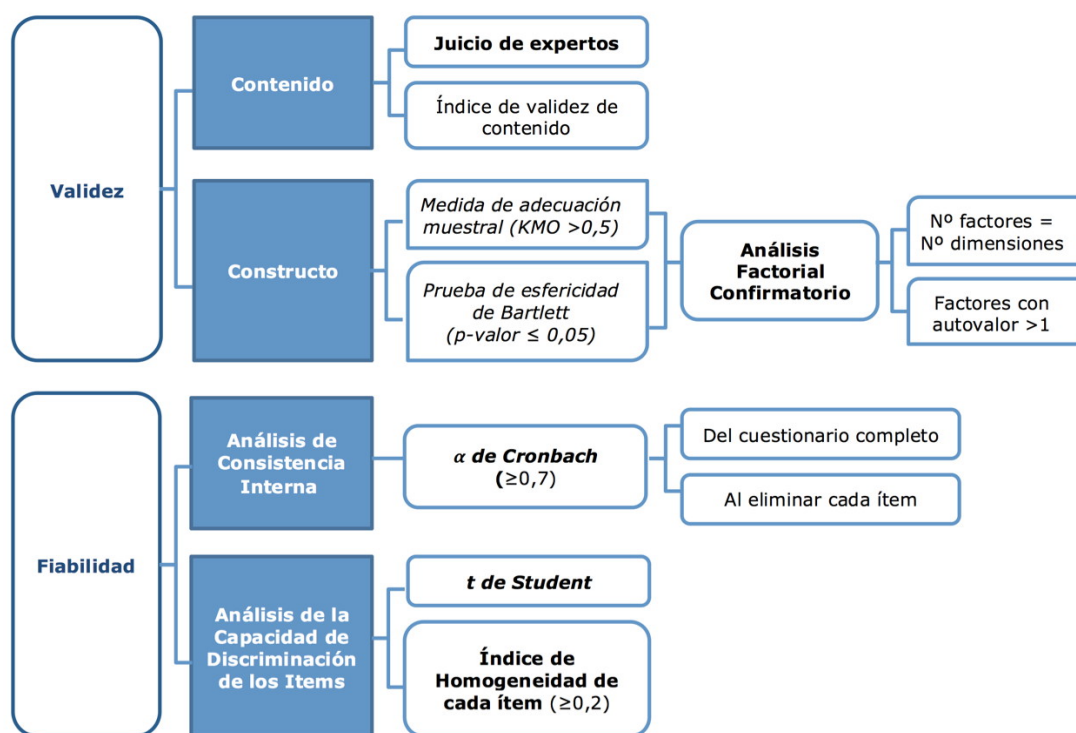


Figura 2: Fases en el proceso de calibración de un cuestionario.

más, es interesante confirmar el resultado anterior mediante la realización de la *prueba de esfericidad de Bartlett* [2], que contrasta la hipótesis nula de que la matriz de correlaciones de las variables es la identidad, en cuyo caso dichas variables no estarían relacionadas. Por tanto, si el nivel de significación del estadístico de Bartlett es mayor que 0.05 no se puede rechazar la hipótesis nula y, por tanto, no tendría sentido realizar el análisis de la estructura dimensional o *análisis factorial* del cuestionario.

El Análisis Factorial (AF) es una técnica de reducción de datos que sirve para encontrar grupos homogéneos de variables a partir de un conjunto numeroso de variables mediante un modelo de regresión múltiple. Fundamentalmente lo que se pretende con el AF es simplificar la información que nos da una matriz de correlaciones para hacerla más fácilmente interpretable y así encontrar respuesta al porqué de que unas variables se relacionen más entre sí y menos con otras. La respuesta que proporciona el AF es porque existen otras variables, *otras dimensiones o factores* que explican por qué unos ítems se relacionan más con unos que con otros. En definitiva, se trata de un análisis de la estructura subyacente a una serie de variables [6].

Una vez confirmada la conveniencia de la realización de dicho análisis, éste suele llevarse a cabo con ayuda de los modelos de *análisis factorial exploratorio* (AFE) o de *análisis factorial confirmatorio* (AFC) [13]. El AFE es recomendable cuando se desea encontrar la mejor solución posible; el AFC,

cuando se pretende confirmar el modelo definido por el investigador o bien el obtenido previamente por los datos. En ambos casos el objetivo del análisis es explicar la varianza común entre las variables con el menor número de factores (*parsimonia*), por lo que intenta que todas las variables entre las que existe una relación o atributo común se agrupen o *saturen* en un mismo factor o dimensión. De esta manera, se podrá afirmar que determinados ítems se explican mejor desde una dimensión que desde otra. Por tanto, si hay N factores, se interpreta como que el instrumento original se puede descomponer en N dimensiones (cada uno compuesto por todos los ítems), en las que cada ítem tienen un *peso específico* distinto según sea su relación con el factor. Por tanto, el AF busca los valores de estos pesos de manera que expliquen la varianza presente en las variables originales.

La *comunalidad* de una variable es la proporción de su varianza que puede ser explicada por el modelo factorial obtenido. Cuando su valor se aproxima a 1 indica que la variable queda totalmente explicada por los factores; mientras que si se aproxima a 0, los factores no explicarán nada la variabilidad de las variables. Esto implica que los ítems del cuestionario representados por variables con comunalidades cercanas a 0 se pueden eliminar del mismo. El resultado del análisis factorial implica reducir la dimensionalidad de la matriz de correlaciones a un número menor de factores, sin pérdida del poder informativo.

Los *autovalores* indican la cantidad de varianza total que

está explicada por cada factor, siendo los que tienen un valor mayor que 1 los que se suelen extraer en el AFE. Para el análisis de un cuestionario docente, que el profesor diseña teniendo en cuenta las dimensiones que quiere medir y los ítems que contendrán cada una de ellas, el análisis recomendable es el AFC [13]. En este caso, para evaluar la bondad del modelo⁴ se puede utilizar la prueba χ^2 , aunque por su sensibilidad al tamaño de la muestra se debería complementar con otros índices [11].

Los métodos de extracción de factores pueden no proporcionar matrices de pesos factoriales adecuadas para su interpretación. Los procedimientos de *rotación de factores* buscan factores cuya matriz de pesos factoriales los hagan más fácilmente interpretables sin alterar el modelo [6]. De esta manera, dado que hay más variables que factores comunes, cada factor tendrá una correlación alta con un grupo de variables y baja con el resto de las variables. Se distinguen dos tipos de rotaciones: la rotación *ortogonal*, que se aplica cuando se supone que los factores en la población no están correlacionados entre sí; y la rotación *oblicua*, que se aplica cuando se supone que los factores en la población están fuertemente correlacionados. Por tanto, para facilitar la interpretación del modelo factorial obtenido y lograr una solución más acorde con el modelo previo, se puede realizar una *rotación de la solución* [13]. En general, en el caso de los cuestionarios docentes, se prefiere la oblicua dada la correlación que suele existir entre sus dimensiones [15].

2.2. Fiabilidad

La fiabilidad 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. Una fiabilidad alta indica que el cuestionario 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 evaluación de la fiabilidad de un cuestionario conlleva, entre otros, la realización de los siguientes análisis:

⁴La evaluación de la bondad de un modelo consiste en ver en qué medida se ajustan los datos observados a una distribución teórica o esperada y se suele utilizar el estadístico χ^2 , que mide la distancia entre dos distribuciones; así, cuanto más cercano a 0 esté el valor obtenido, más ajustadas están las distribuciones.

⁵La escala de tipo Likert [22] es una escala psicométrica mediante la que se especifica el nivel de acuerdo o desacuerdo con un ítem. El formato de las respuestas de un típico elemento de Likert con 5 niveles de respuesta sería: (1: Totalmente en desacuerdo; 2: En desacuerdo; 3: Ni de acuerdo ni en desacuerdo; 4: De acuerdo; 5: Totalmente de acuerdo).

⁶Un contraste de hipótesis es un procedimiento para decidir si una propiedad que se supone en una población estadística es compatible con lo observado en una muestra de dicha población. Para ello se define una hipótesis determinada, llamada hipótesis nula y se calcula la distancia que hay entre ella y la información que proporciona la muestra del estudio mediante el cálculo de un estadístico. En este contexto, el p-valor se define como la probabilidad de obtener un resultado al menos tan extremo como el que realmente se ha obtenido para el estadístico calculado, de manera que cuanto menor sea el p-valor mayor es la incompatibilidad de la muestra con la hipótesis nula.

⁷La prueba t de Student es cualquier prueba en la que el estadístico utilizado tiene una distribución t de Student si la hipótesis nula es cierta. Se aplica cuando la población estudiada sigue una distribución normal pero el tamaño muestral es demasiado pequeño como para que el estadístico en el que está basada la inferencia esté normalmente distribuido, utilizándose una estimación de la desviación típica en lugar del valor real.

Análisis de consistencia interna, con el objeto de dotar de significación a las preguntas del cuestionario. Para ello se suele calcular el coeficiente *alfa de Cronbach* [5], que está basado en la correlación interelementos promedio y asume que los ítems, medidos en escala tipo Likert,⁵ miden un mismo constructo y que están altamente correlacionados [36]. Los valores del coeficiente alfa de Cronbach oscilan entre 0 y 1, considerando como criterio general un coeficiente *aceptable* cuando su valor es igual o superior a 0.7 [17], ya que este estadístico no va acompañado de ningún p-valor.⁶ Además, se aconseja evaluar el valor del coeficiente alfa de Cronbach al eliminar del cuestionario cada uno de los ítems que lo componen, ya que se puede prescindir de aquellos en los al ser eliminados hacen que el valor del coeficiente aumente.

Análisis de la capacidad de discriminación de los ítems de modo que se refuerce el carácter unidimensional de la prueba. Se puede utilizar la prueba *t de Student*,⁷ para contrastar la hipótesis nula que indica la no existencia de diferencias entre las medias de los grupos establecidos, así como el *índice de homogeneidad* de cada ítem, esto es el coeficiente de correlación de Pearson entre la puntuación en el ítem y la suma de las puntuaciones en los restantes ítems. El índice de homogeneidad de un ítem nos va a informar del grado en que dicho ítem está midiendo lo mismo que la prueba globalmente; es decir, del grado en que contribuye a la consistencia interna del test. Los ítems con bajos índices de homogeneidad miden algo diferente a lo que refleja la prueba en su conjunto, por lo que se pueden eliminar del cuestionario. Normalmente, se eliminan aquellos cuyo índice de homogeneidad es menor que 0,2 [16].

3. Calidad del cuestionario CoMaR

En las siguientes secciones se describe en detalle cómo hemos aplicado el método a la evaluación de la calidad del cuestionario CoMaR [20] y las mejoras que este proceso ha supuesto en la definición de su contenido final.

3.1. Características externas

El cuestionario lo completaron 77 alumnos de segundo curso del Grado en Informática de la Escuela Superior de Informática de Ciudad Real, perteneciente a la Universidad de

Castilla-La Mancha. La actividad se desarrolló durante la primera clase de la asignatura obligatoria de segundo curso Metodología de la Programación, que se imparte en el segundo cuatrimestre, por lo que los alumnos ya han cursado previamente las asignaturas de Fundamentos de Programación I y Fundamentos de Programación II, de primer curso, y la de Estructuras de Datos, de segundo curso.

Nuestra experiencia como profesores en la enseñanza de la programación nos ha servido para intuir que las principales dificultades de los estudiantes con la recursividad tiene que ver con dos factores: las dificultades cognitivas inherentes al concepto recursivo y la predisposición del aprendiz frente a ella. Por ello, el contenido del cuestionario se diseñó con el objetivo de buscar alguna relación entre las principales *dificultades* que conlleva el aprendizaje de la recursividad, con el *conocimiento previo* adquirido por el estudiante sobre la base teórica de la recursividad y con su *actitud personal*. Teniendo en cuenta, además, que entre las dificultades podemos distinguir las relacionadas con el *análisis* del problema, las que tienen que ver con el *diseño* del algoritmo y las relacionadas con la ejecución o seguimiento de la *traza* del programa recursivo. Con el fin de concretar la generalidad de estas 6 dimensiones, se definieron 27 ítems agrupados tal y como se presenta en el cuadro 1.

Las respuestas se recogen mediante escalas de Likert, con un puntaje mínimo de 1 y un máximo de 5, donde las puntuaciones bajas indican discrepancia y las altas corresponden a mayor acuerdo con la sentencia enunciada en la pregunta. Hemos elegido esta escala porque los expertos recomiendan utilizar respuestas graduadas siempre que sea posible [13]. Además, este tipo de escalas son fáciles de utilizar e interpretar por el profesor y muy intuitivas para el alumnado. Incluso el análisis factorial que vamos a realizar para analizar la fiabilidad del cuestionario es más eficaz cuando la escala de Likert es de al menos 5 categorías [13, 15]. Las características psicométricas del cuestionario CoMaR se obtuvieron mediante distintos tipos de análisis estadísticos realizados con la ayuda del programa IBM SPSS Statistics, versión 19 [17].

3.2. Análisis de validez

El grupo de jueces para validar el contenido del cuestionario CoMaR estuvo formado por 9 expertos en el tema de la recursividad, de los que 6 son profesores de la Universidad de Castilla-La Mancha, 2 pertenecen a la Universidad Rey Juan Carlos de Madrid y 1 a la Universidad de Granada; todos ellos dedicados a la docencia de la programación y con experiencia contrastada en la enseñanza de la recursividad. A todos se les proporcionó el mismo documento en el que se especificaba claramente el objetivo del cuestionario, su contenido (cuadro 1), así como la forma en la que debían aportar su valoración sobre las dimensiones definidas, los ítems asociados a cada una de ellas y la escala de valoración de los mismos. El 90 % de los jueces coincidió en mantener las 6 dimensiones y los 27 ítems originales, así como la escala de Likert de valoración, por lo que no hemos considerado necesario calcular el

índice de validez del contenido.

A continuación, se procedió a la realización del análisis factorial del cuestionario CoMaR con los siguientes objetivos: por un lado, comprobar si los ítems están bien agrupados en las 6 dimensiones que hemos definido inicialmente; por otro, si dichas dimensiones son suficientes para explicar los resultados que proporcionan las 27 preguntas que contiene; y por último, intentar reducir la matriz de correlaciones a otra equivalente pero con un número menor de factores y sin pérdida del poder informativo, ya que el análisis de dicha matriz es tremendamente complejo debido a su tamaño.

Antes de comenzar con el análisis factorial, confirmamos que merece la pena realizarlo mediante el cálculo de la medida de adecuación muestral KMO, para la que obtuvimos un valor de 0,656 ($> 0,5$), y de la *prueba de esfericidad de Bartlett*, cuyo valor fue 0. Así pues, decidimos realizar un AFC definiendo 6 factores de extracción, correspondientes a las 6 dimensiones del cuestionario. En la segunda columna del cuadro 2 observamos que los 6 primeros factores tienen autovalores mayores que 1 y explican un 64 % de la varianza total de los datos originales. Sin embargo, el método de extracción elegido no converge con ninguno de los métodos de extracción disponibles para el análisis factorial en SPSS. Según refleja el cuadro 2, se deberían extraer 8 factores (los que tienen autovalores mayores que 1), dos más de los que hemos definido inicialmente, por lo que decidimos analizar la fiabilidad del cuestionario por si este análisis sugiere la eliminación de algún ítem y, por ende, de algún factor.

3.3. Análisis de la fiabilidad

El índice de fiabilidad del cuestionario completo que se obtiene mediante el cálculo del estadístico alpha de Cronbach es de 0,708, que se considera aceptable. Este valor aumenta sensiblemente si se eliminan los ítems P1, P4, P5 y P17, como ilustra el cuadro 3. Si tenemos en cuenta además aquellos cuyo índice de homogeneidad es menor que 0,2, se puede prescindir también de las preguntas P3, P4, P7, P8, P10 y P25. Al eliminar los 10 ítems indicados volvimos a realizar un análisis de fiabilidad para los datos resultantes, obteniendo un valor para el alpha de Cronbach de 0,784, más alto que en el caso original. Ahora el análisis refleja que tres de los ítems, P2, P6 y P9, se pueden excluir puesto que sus *índices de homogeneidad* están muy cercanos a 0 y al eliminarlos el índice de fiabilidad del cuestionario aumenta (cuadro 4).

El índice de fiabilidad del cuestionario reducido a los 14 ítems restantes es de 0,825, un valor que se considera aceptable [17]. Además, los índices de homogeneidad de todos los ítems, enumerados en el cuadro 5, son mayores o iguales a 0,2. Por tanto, podemos considerar que el cuestionario resultante presenta una buena consistencia interna.

Conocimiento previo
P1. Conozco los principios teóricos de la recursividad
P2. Comprendo los principios teóricos de la recursividad
P3. Conozco el funcionamiento de la recursividad
P4. Comprendo el funcionamiento de la recursividad
P5. Conozco el principio de inducción matemática
P6. Comprendo el principio de inducción matemática
Actitud
P7. Me gusta programar
P8. En general, me gusta la recursividad
Utilidad percibida
P9. Considero útil la recursividad para la solución de problemas complejos
P10. Valoro la utilidad de la recursividad en programación
Dificultades Diseño
P11. Me cuesta trabajo pensar de forma declarativa
P12. Me cuesta trabajo diseñar la solución recursiva de un problema
P13. Me cuesta trabajo definir el caso base
P14. Me cuesta trabajo “creerme” el caso $n - 1$
P15. Me cuesta trabajo definir el caso general
Dificultades Análisis
P16. Me cuesta trabajo analizar (tratar de entender) el funcionamiento de programas recursivos
P17. Me cuesta trabajo entender el caso base
P18. Me cuesta trabajo entender el caso $n - 1$
P19. Me cuesta trabajo entender el caso general
Dificultades Ejecución
P20. La llegada al caso base desde el caso general
P21. Acceso y modificación de parámetros y variables locales
P22. Acceso y modificación de las variables globales
P23. Llamadas activas
P24. Recuperación de los valores de variables y parámetros en la vuelta atrás de la llamada activa
P25. Paso de objetos como parámetros
P26. Funcionamiento del programa cuando existen varias llamadas recursivas
P27. Seguimiento de la traza de un algoritmo recursivo

Cuadro 1: Contenido del cuestionario CoMaR tal y como fue diseñado inicialmente.

Factor	Autovalor	% de la varianza	% acumulado
1	6,175	22,872	22,872
2	3,132	11,600	34,472
3	2,389	8,847	43,318
4	2,252	8,339	51,658
5	1,842	6,824	58,482
6	1,470	5,444	63,926
7	1,245	4,613	68,538
8	1,046	3,873	72,411
9	0,916	3,393	75,804
10	0,819	3,034	78,839
11	0,723	2,676	81,515
12	0,629	2,330	83,845
13	0,600	2,223	86,068
14	0,543	2,011	88,079
15	0,485	1,795	89,874
16	0,445	1,648	91,522
17	0,380	1,406	92,929
18	0,356	1,317	94,246
19	0,290	1,074	95,319
20	0,250	0,926	96,245
21	0,202	0,747	96,992
22	0,190	0,704	97,697
23	0,185	0,687	98,383
24	0,161	0,596	98,979
25	0,114	0,424	99,403
26	0,089	0,329	99,732
27	0,072	0,268	100,00

Cuadro 2: Varianza total explicada por los autovalores iniciales obtenidos mediante el método de extracción de máxima verosimilitud. Obsérvese que para el cálculo de los autovalores se define un factor por cada uno de los ítems que componen el cuestionario de tal manera que los 27 ítems en su conjunto explican, lógicamente, el 100 % de la varianza total del modelo.

Ítem	Índice de homogeneidad	α de Cronbach al eliminar el ítem
P1	0,123	0,724
P2	0,258	0,707
P3	0,049	0,699
P4	-0,137	0,711
P5	0,188	0,725
P6	0,255	0,700
P7	-0,148	0,705
P8	0,002	0,697
P9	0,258	0,700
P10	0,195	0,678
P11	0,252	0,696
P12	0,243	0,682
P13	0,183	0,682
P14	0,276	0,705
P15	0,233	0,699
P16	0,496	0,699
P17	0,306	0,719
P18	0,479	0,699
P19	0,497	0,703
P20	0,284	0,697
P21	0,404	0,687
P22	0,397	0,689
P23	0,350	0,693
P24	0,364	0,689
P25	0,150	0,707
P26	0,212	0,703
P27	0,350	0,692

Cuadro 3: Valores de los índices de homogeneidad de cada ítem y del índice de fiabilidad del cuestionario al eliminar cada ítem.

Ítem	Índice de homogeneidad	α de Cronbach al eliminar el ítem
P2	0,048	0,793
P6	0,036	0,799
P9	0,042	0,796
P11	0,238	0,783
P12	0,366	0,774
P14	0,366	0,773
P15	0,344	0,775
P16	0,596	0,754
P18	0,562	0,759
P19	0,625	0,755
P20	0,275	0,780
P21	0,457	0,766
P22	0,492	0,765
P23	0,449	0,769
P24	0,518	0,761
P26	0,427	0,769
P27	0,512	0,762

Cuadro 4: Valores de los índices de homogeneidad de cada ítem y del índice de fiabilidad del cuestionario reducido al eliminar cada pregunta.

Ítem	Índice de homogeneidad	α de Cronbach al eliminar el ítem
P11	0,200	0,832
P12	0,386	0,819
P14	0,392	0,819
P15	0,394	0,818
P16	0,607	0,802
P18	0,585	0,805
P19	0,688	0,799
P20	0,309	0,824
P21	0,422	0,817
P22	0,476	0,813
P23	0,471	0,814
P24	0,504	0,811
P26	0,448	0,816
P27	0,521	0,810

Cuadro 5: Valores de los índices de homogeneidad de cada ítem y del índice de fiabilidad del cuestionario después de su segunda simplificación a 14 preguntas.

3.4. Análisis factorial exploratorio del cuestionario reducido

El siguiente paso es el de determinar el número óptimo de factores o dimensiones del cuestionario reducido mediante un análisis factorial exploratorio, ya que al haberse reducido tanto el número de ítems (prácticamente se han eliminado la mitad) el modelo original de 6 dimensiones ha dejado de tener sentido. El valor de la medida de adecuación muestral KMO ahora es de 0,741 ($> 0,5$), y mayor que el del cuestionario completo original (véase sección 3.2). Además, el valor del nivel de significación de la prueba de esfericidad de Bartlett es 0. Por tanto, tiene sentido realizar un análisis factorial para el que utilizaremos el método de *máxima verosimilitud* para la extracción de factores. Elegimos este método de extracción porque el programa SPSS incluye el cálculo del *índice de la bondad del ajuste*. En este contexto, la hipótesis nula considera que la matriz de correlación entre las variables observadas de la muestra puede ser reproducida exactamente por el modelo factorial en la población. El rechazo de dicha hipótesis debería interpretarse como que el modelo propuesto no se cumple exactamente en la población. Puesto que pretendemos encontrar el número de factores que mejor se adapte a nuestro modelo, la extracción no se ha realizado para un número fijo de factores, como en el caso inicial (sección 3.2), sino para todos aquellos cuyos autovalores sean mayor que 1, siendo sólo 3 en este caso los que explican el 56 % de la varianza total. Además, con el fin de identificar mejor las saturaciones de cada ítem respecto a cada factor, hemos aplicado un método de rotación oblicua (Oblimin normal). Con estos datos, el índice de bondad del ajuste o p-valor, calculado mediante la prueba de χ^2 , es de 0,435. No podemos rechazar la hipótesis nula y asumimos un modelo en base a los 3 factores obtenidos. El cuadro 6 contiene las saturaciones de cada una de las

preguntas en cada factor antes y después de la rotación donde se observa que las comunalidades son todas mayores que 0.

4. Discusión de los resultados

Los datos del cuestionario ofrecen un alto índice de fiabilidad cuando se eliminan las preguntas relacionadas con el *conocimiento previo* (P1 a P6), la *actitud* del alumno (P7 y P8) y la *utilidad* percibida (P9 y P10). Esto se interpreta como que las respuestas dadas por los alumnos a estas preguntas no aportan información en relación a la confianza que se concede a los datos que se obtienen con el cuestionario, por lo que se puede prescindir de ellas a la hora de medir las dificultades del alumno en el aprendizaje de la recursividad. Y confirma los resultados obtenidos en estudios previos [20] en los que no se encontró relación significativa entre dichos factores y los demás. Para nosotros tiene sentido que estos tres factores no aparezcan como factores a tener en cuenta en el aprendizaje de la recursividad porque en el momento en el que se realizó la experiencia, el comienzo de una asignatura de segundo semestre, los alumnos tienen, en general, una actitud positiva hacia las nuevas asignaturas que comienzan a estudiar y perciben la utilidad de la recursividad pues, a través de la guía docente y la presentación por parte del profesorado de la asignatura han asimilado que es una herramienta necesaria para tener éxito en su aprendizaje. Respecto al conocimiento previo, los alumnos creen que tienen los conocimientos teóricos necesarios para el manejo de la recursividad por lo que no les influye en su percepción sobre las dificultades de la recursividad. Seguramente es porque aunque tienen una idea de lo que es y de cómo funciona, no han practicado con ella lo suficiente para darse cuenta de sus propias limitaciones en cuanto al conocimiento necesario.

Por otra parte, la eliminación de las preguntas P13 y P17,

Ítem	Modelo Inicial			Después de la rotación		
	F1	F2	F3	F1	F2	F3
P11	0,162	0,236	0,051	0,019	0,249	0,189
P12	0,441	0,042	-0,208	0,397	0,401	0,079
P14	0,419	0,410	-0,362	0,230	0,670	0,040
P15	0,400	0,192	-0,261	0,300	0,485	0,054
P16	0,607	0,207	-0,135	0,449	0,599	0,270
P18	0,518	0,724	-0,270	0,151	0,912	0,250
P19	0,643	0,477	0,002	0,329	0,755	0,470
P20	0,276	0,135	-0,038	0,180	0,291	0,152
P21	0,400	0,247	0,700	0,105	0,217	0,841
P22	0,467	0,187	0,681	0,195	0,231	0,845
P23	0,453	0,132	0,550	0,232	0,228	0,718
P24	0,577	-0,023	0,202	0,473	0,325	0,464
P26	0,809	-0,505	-0,105	0,956	0,275	0,214
P27	0,616	-0,014	0,008	0,537	0,419	0,331

Cuadro 6: Estructura factorial antes y después de la rotación.

que están relacionadas con la definición y comprensión del caso base, también tiene sentido pues refleja el hecho de que los aspectos relacionados con el caso base no suponen ninguna dificultad para el estudiante, por lo que se pueden obviar dichas preguntas.

De modo similar, también tiene sentido la eliminación del cuestionario de la pregunta 25 porque se ha comprobado empíricamente [20] que el paso de objetos como parámetros no les supone ninguna dificultad específica, ya que éstos actúan como variables globales y la pregunta P22 ya tiene en cuenta ese hecho.

Por último, en función del modelo factorial obtenido, la estructura del cuestionario se puede describir en base a tres factores o dimensiones: las *conceptuales*, que requieren un mayor nivel de abstracción; las relacionadas con la *gestión de las variables*, conforme se van generando las distintas llamadas recursivas; y las relacionadas con lo que implica la *vuelta atrás*.

En consecuencia, el cuestionario CoMaR* de 14 preguntas, descrito en el cuadro 7, ofrece mejores propiedades psicométricas que el cuestionario original, tanto a nivel general como para cada una de sus dimensiones específicas. Esto sugiere que, de cara a un análisis estadístico posterior, los datos proporcionados por estas 14 preguntas tienen más validez y fiabilidad que los que ofrecen las 27 preguntas del cuestionario original.

5. Conclusiones

En este trabajo hemos sometido a estudio la calidad del cuestionario CoMaR, utilizado en trabajos previos para identificar las dificultades de los alumnos durante el aprendizaje de la recursividad. Como resultado final, se ha obtenido un cuestionario simplificado, basado en un modelo de tres factores y

con un total de 14 preguntas, 13 menos de las que contenía el cuestionario original, y que explican el 54 % de la varianza total de los datos y con una buena consistencia interna.

Los resultados de este estudio sugieren que es necesario que el investigador posea conocimientos de estadística para garantizar la calidad de los análisis y la validez de los resultados obtenidos. Por tanto, creemos que este trabajo aporta a la comunidad educativa una serie de pautas para analizar la calidad de un cuestionario docente, como paso previo a la realización de cualquier tipo de análisis estadístico de los datos recogidos por él y mediante el que se quieran extraer conclusiones válidas.

Además, este trabajo ha identificado una escala breve (CoMaR*) que permitirá valorar la magnitud del problema de la recursividad, de forma rápida, de bajo coste y en la mayor parte de contextos posibles. De hecho, en el futuro nos planteamos utilizar el cuestionario reducido en lugar del planteado inicialmente.

Sin embargo, entre las limitaciones del estudio es importante mencionar dos de ellas: la primera, que el uso del paquete estadístico SPSS resulta insuficiente a la hora de realizar un análisis factorial completo de un instrumento de medida porque no proporciona métodos suficientes para el cálculo del índice de bondad del ajuste del modelo factorial. Esta carencia se puede solventar utilizando programas más potentes, como por ejemplo R (<https://www.r-project.org>) pero ello exige un esfuerzo mayor al docente pues para su manejo es necesario disponer, no sólo de conocimientos sobre estadística, sino también de programación, un hándicap importante de cara a su uso para experimentos docentes ajenos al contexto de la estadística.

La segunda limitación está relacionada con la muestra del estudio, porque aunque su tamaño es aceptable [29], no es representativa de todos los estudiantes de programación, por lo que las conclusiones obtenidas son sólo significativas para la

Dificultades Conceptuales
PR1. Me cuesta trabajo pensar de forma declarativa
PR2. Me cuesta trabajo diseñar la solución recursiva de un problema
PR3. Me cuesta trabajo “creerme” el caso $n - 1$
PR4. Me cuesta trabajo definir el caso general
PR5. Me cuesta trabajo analizar (tratar de entender) el funcionamiento de programas recursivos
PR6. Me cuesta trabajo entender el caso $n - 1$
PR7. Me cuesta trabajo entender el caso general
PR8. La llegada al caso base desde el caso general
Dificultades Gestión de variables
PR9. Acceso y modificación de parámetros y variables locales
PR10. Acceso y modificación de las variables globales
PR11. Llamadas activas
Dificultades con la Vuelta Atrás
PR12. Recuperación de los valores de variables y parámetros en la vuelta atrás de la llamada activa
PR13. Funcionamiento del programa cuando existen varias llamadas recursivas
PR14. Seguimiento de la traza de un algoritmo recursivo

Cuadro 7: Cuestionario CoMaR*, después del análisis de la calidad del cuestionario inicial, etiquetando de nuevo los ítems que lo integran.

población del estudio, que comprende los estudiantes del Grado de Informática de la Escuela Superior de Informática de la UCLM.

No obstante, nuestro trabajo puede ser considerado como un *estudio piloto* que se puede ampliar mediante su replicación con alumnos de otros centros y mejorar así las propiedades del cuestionario; en concreto la del análisis factorial, cuya estabilidad aumenta a medida que se incrementa el tamaño de la muestra [13].

Referencias

- [1] American Educational Research Association; American Psychological Association; National Council on Measurement in Education (AERA/APA/NCME). *Standards for educational and psychological testing*. Washington, DC, 1999.
- [2] M.S. Bartlett. *Tests of significance in factor analysis*. British Journal of Statistical Psychology, vol. 3, núm. 2, pp.77–85. Junio de 1950.
- [3] Hugh Burkhardt y Alan H. Schoenfeld. *Improving Educational Research: Toward a More Useful, More Influential, and Better-Funded Enterprise*. Educational Researcher, vol. 32, núm. 9, pp. 3–14. Diciembre de 2003.
- [4] Alberto Cernuda del Río y Miguel Riesco Albizu. *Bolonia 15 años después*. En Actas de las XXI Jornadas de la Enseñanza Universitaria de la Informática, Jenui 2015, pp. 333–340. Andorra la Vella, julio de 2015.
- [5] Lee J. Cronbach. *Coefficient alpha and internal structure of tests*. Psychometrika, vol. 16, núm. 3, pp. 297–334. Septiembre de 1951.
- [6] Santiago de la Fuente Fernández. *Análisis Factorial*. 2011. Disponible en <http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/MULTIVARIANTE/FACTORIAL/analisis-factorial.pdf>. Última consulta, diciembre de 2015.
- [7] Declaración conjunta de los Ministros Europeos de Educación. *Declaración de Bolonia*. Junio de 1999. Disponible en http://www.educacion.gob.es/boloniaensecundaria/img/Declaracion_Bolonia.pdf. Última consulta, diciembre de 2015.
- [8] Elisabet Díaz Costa, Antonio Fernández-Cano, Tarik Faouzi y Carlos Felipe Henríquez. *Validación del constructo subyacente en una escala de evaluación del impacto de la investigación educativa sobre la práctica docente mediante análisis factorial confirmatorio*. Revista de Investigación Educativa, vol. 33, núm. 1, pp. 47–63. 2015.
- [9] Jazmine Escobar-Pérez y Ángela Cuervo-Martínez. *Validez de contenido y juicio de expertos: una aproximación a su utilización*. Avances en Medición, vol. 6, núm. 1, pp. 27–36. 2008.

- [10] Juan M. Escudero Muñoz. *Compartir propósitos y responsabilidades para una mejora democrática de la educación*. Revista de Educación, núm. 339, pp. 19–41. 2006.
- [11] María Fernández, Juan L. Benítez, M^a Carmen Pichardo, Eduardo Fernández, Fernando Justicia, Trinidad García, Ana García-Berbén, Ana Justicia y Guadalupe Alba. *Análisis factorial confirmatorio de las subescalas del PKBS-2 para la evaluación de las habilidades sociales y los problemas de conducta en educación infantil*. Electronic Journal of Research in Educational Psychology, vol. 8, núm. 3, pp. 1229–1252. Diciembre de 2010.
- [12] Antonio Fernández Cano. *Valoración del impacto de la investigación educativa sobre la práctica docente*. Revista de Educación, núm. 324, pp. 155–170. 2001.
- [13] Pere Joan Ferrando y Cristina Anguiano-Carrasco. *El análisis factorial como técnica de investigación en Psicología*. Papeles del Psicólogo, vol. 31, núm. 1, pp. 18–33. Enero de 2010.
- [14] Peter W. Frey, Dale W. Leonard y William W. Beatty. *Student Ratings of Instruction: Validation Research*. American Educational Research Journal, vol. 12, núm. 4, pp. 435–444. Septiembre de 1975.
- [15] Dolores Frías-Navarro y Marcos Pascual Soler. *Prácticas del análisis factorial exploratorio (AFE) en la investigación sobre conducta del consumidor y marketing*. Suma Psicológica, vol. 19, núm. 1, pp. 47–58. Junio de 2012.
- [16] M^a Isabel García Barbero, Enrique Vila Abad, y Juan Carlos Suárez Falcón. *Psicometría*. Madrid: UNED, 2006.
- [17] Darren George y Paul Mallery. *SPSS for Windows Step by Step. A Simple Guide and Reference*. 4^a ed. Allyn & Bacon, Boston 2002.
- [18] Carl F. Kaestle. *The Awful Reputation of Education Research*. Educational Researcher, vol. 22, núm. 1, pp. 23+26–31. Enero–febrero de 1993.
- [19] Henry F. Kaiser. *An Index of Factorial Simplicity*. Psychometrika, vol. 39, núm. 1, pp. 31–36. Marzo de 1974.
- [20] Carmen Lacave, Ana Isabel Molina y Juan Giralt. *Identificando algunas causas del fracaso en el aprendizaje de la recursividad. Análisis experimental en las asignaturas de programación*. En Actas de las XIX Jornadas de la Enseñanza Universitaria de la Informática, Jenui 2013, pp. 225–232. Castellón, julio de 2013.
- [21] C. H. Lawshe. *A quantitative approach to content validity*. Personnel Psychology, vol. 28, núm.4, pp. 563–575. Diciembre de 1975.
- [22] Rensis Likert. *A technique for the measurement of attitudes*. Archives of Psychology, vol. 22, núm. 140, pp. 1–55. Junio de 1932.
- [23] Irene Lucero y Susana Meza. *Validación de instrumentos para medir conocimientos*. UNNE, Departamento de Física - Facultad de CC. Exactas y Naturales y Agrimensura, Argentina. 2002. Disponible en <http://www1.unne.edu.ar/cyt/2002/09-Educacion/D-027.pdf>. Última consulta, diciembre de 2015.
- [24] José Francisco Lukas, Karlos Santiago, Juan Etxeberria y Luis Lizasoain. *Adaptación al Espacio Europeo de Educación Superior de un cuestionario de opinión del alumnado sobre la docencia de su profesorado*. Relieve, vol. 20, núm. 1. 2014.
- [25] H. W. Marsh. *SEQ: A reliable, valid and useful instrument for collecting students' evaluations of university teaching*. British Journal of Educational Psychology, vol. 52, núm. 1, pp. 77–95. Febrero de 1982.
- [26] Cástor Méndez Paz. *La implantación del sistema de créditos europeo como una oportunidad para la innovación y mejora de los procedimientos de enseñanza-aprendizaje en la Universidad*. Revista Española de Pedagogía, vol. 63, núm. 230, pp. 43–62. Enero–abril de 2005.
- [27] Ministerio de Educación, Cultura y Deporte. *La integración del Sistema Universitario Español en el Espacio Europeo de Enseñanza Superior. Documento-Marco*. Febrero de 2003. Disponible en <http://tecnologiaedu.us.es/mec2011/html/mas/2/21/7.pdf>. Última consulta, diciembre de 2015.
- [28] Pedro Morales Vallejo. *La fiabilidad de los tests y las escalas*. Universidad Pontificia Comillas, Madrid. Facultad de Ciencias Humanas y Sociales. Septiembre de 2007. Disponible en <http://web.upcomillas.es/personal/peter/estadisticabasica/Fiabilidad.pdf>. Última consulta, diciembre de 2015.
- [29] Pedro Morales Vallejo. *Guía para construir cuestionarios y escalas de actitudes*. Universidad Pontificia Comillas, Madrid. Septiembre de 2011. Disponible en <http://web.upcomillas.es/personal/peter/otrosdocumentos/Guiaparaconstruיריםcalasdeactitudes.pdf>. Última consulta, diciembre de 2015.
- [30] Gerardo Prieto y Ana R. Delgado. *Fiabilidad y Validez*. Papeles del Psicólogo, vol. 31, núm. 1, pp. 67–74. Enero de 2010.
- [31] Carlos Ruiz Bolívar. *Validez*. UPEL/PIDE. Disponible en <http://investigacion.upeu.edu.pe/>

images/7/74/Validez.pdf. Última consulta, diciembre de 2015.

- [32] Raja Sooriamurthi. *Problems in Comprehending Recursion and Suggested Solutions*. En Proceedings of the 6th annual conference on Innovation and Technology in Computer Science Education, ITiCSE. pp. 25–28. Canterbury, UK, junio de 2001.
- [33] Javier Tourón. *La validación de constructo: su aplicación al CEED (cuestionario para la evaluación de la eficacia docente)*. Bordón, vol. 41, núm. 4, pp. 735-756. 1989.
- [34] J. Ángel Velázquez-Iturbide. *Recursion in Gradual Steps (Is Recursion Really that Difficult?)*. En Proceedings of the thirty-first SIGCSE technical symposium on Computer Science Education, SIGCSE, vol. 32, núm. 1, pp. 310–314. Marzo de 2000.
- [35] M^a Victoria Verdugo Matés y M^a Isabel Cal Bouzada. *Valoración de la enseñanza: SEEQ*. Revista de Formación e Innovación Educativa Universitaria, vol. 3, núm. 4, pp. 182–193. 2010.
- [36] Susan Welch y John Comer. *Quantitative Methods for Public Administration: Techniques and Applications*. Fort Worth: Harcourt Brace College Publishers, 1988.
- [37] Miguel A. Zabalza. *Innovación en la Enseñanza Universitaria* Contextos Educativos, núm. 6–7, pp. 113–136. 2003-2004.

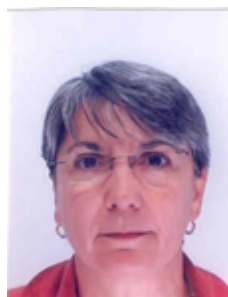


Carmen Lacave Rodero es Licenciada en CC. Matemáticas (1990) por la UCM y Doctora en Ciencias (2003) por la UNED. En la actualidad es Profesora Titular del área de Lenguajes y Sistemas Informáticos de la Universidad de Castilla-La Mancha, adscrita a la Escuela Superior de Informática.

Pertenece al grupo de investigación CHICO de la UCLM, su trabajo se centra en la aplicación de las Tecnologías de la Información y de la Inteligencia Artificial a la educación.



Ana Isabel Molina Díaz es Ingeniera en Informática (2002) y Doctora (2007) por la Universidad de Castilla-La Mancha (España). Se encuentra adscrita a la Escuela Superior de Informática de la Universidad de Castilla-La Mancha. Desde el año 2003 es miembro del grupo de investigación CHICO (Computer-Human Interaction and Collaboration) de la Universidad de Castilla-La Mancha. Además de la enseñanza, sus principales áreas de interés son las Tecnologías de la Información aplicadas a la educación, el diseño y especificación de interfaces colaborativas y la evaluación de recursos educativos mediante técnicas de seguimiento ocular.



Mercedes Fernández es Licenciada en CC. Matemáticas por la Universidad de Granada y Profesora Titular de Escuela Universitaria del área de Estadística e Investigación Operativa del Departamento de Matemáticas de la Universidad de Castilla-La Mancha, adscrita a la Escuela Superior de Informática. Su investigación se enmarca dentro del Grupo de Diseño y Análisis de Experimentos de la UCLM, centrándose en el diseño óptimo de experimentos.



Miguel Ángel Redondo es Doctor en Ingeniería Informática por la Universidad de Castilla - La Mancha (2002), Licenciado en Informática por la Universidad de Granada (1997), Ingeniero Técnico en Informática de Sistemas por la Universidad de Castilla - La Mancha (1995). Actualmente es Profesor Titular del área de Lenguajes y Sistemas Informáticos de la Universidad de Castilla - La Mancha, adscrito a la Escuela Superior de Informática. A nivel de investigación, su trabajo se centra en la innovación y aplicación de técnicas de Ingeniería del Software al desarrollo de sistemas de e-Learning y al desarrollo de sistemas avanzados de Interacción Persona-Computador.



© 2016 C. Lacave, A.I. Molina, M. Fernández, M.A. Redondo. Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional que permite copiar, distribuir y comunicar públicamente la obra en cualquier medio, sólido o electrónico, siempre que se acrediten a los autores y fuentes originales y no se haga un uso comercial.