

# Análisis de la fiabilidad y validez de un cuestionario docente

Carmen Lacave Rodero<sup>1</sup>, Ana Isabel Molina Díaz<sup>1</sup>, Mercedes Fernández Guerrero<sup>2</sup>, Miguel Ángel Redondo Duque<sup>1</sup>

<sup>1</sup> Dpto. de Tecnologías y Sistemas de la Información, UCLM - <sup>2</sup> Departamento de Matemáticas, UCLM  
{carmen.lacave, anaisabel.molina, mercedes.fernandez, 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 vale con 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 una edición anterior de las JENUÍ se presentó un cuestionario para medir las dificultades de los estudiantes durante el aprendizaje de la recursividad. En este trabajo describimos el análisis de la validez y fiabilidad de dicho cuestionario. Las conclusiones de este estudio permiten 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.

## Abstract

At present, questionnaires are widely used by teachers to assess different aspects related to teaching, since statistical analysis of the data collected through them allow teachers to infer conclusions in a simple and rigorous way. However, this process is often overlooked because to ensure the usefulness and meaning of the results obtained with it, is not enough the simply presentation of a list of questions designed *ad hoc* by teachers, but the questionnaire must be well designed according to terms of quality. The main quality criteria

required by assessment instruments are reliability and validity. In a previous edition of JENUÍ a questionnaire to measure students' difficulties in learning recursion was presented. In this paper we describe the analysis of the validity and reliability of such questionnaire. The findings of this study allow us to provide the educational community a rapid implementation tool to assess the problem of recursion. In addition, the method described along the article can be easily extrapolated and we believe it can serve as a guidance for the validation of any other educational questionnaire.

## Palabras clave

Cuestionario, calidad, fiabilidad, validez, recursividad.

## 1. Introducción

El método más sencillo y rápido que se suele utilizar para evaluar distintos aspectos relacionados con la docencia es el uso de *cuestionarios*<sup>1</sup>. 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 estándar de calidad. Esta tarea no es sencilla pues construir un cuestionario técnicamente bien hecho conlleva, en sí mismo, una investigación [1]. Por tanto, el uso de cuestionarios implica necesariamente el control de la bondad del mismo mediante el estudio de su *fiabilidad* y *validez*. 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” no se puede garantizar el éxito del diagnóstico, ya que el

<sup>1</sup> A 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 en base a alguna escala.

tratamiento estadístico no conseguirá transformar datos de mala calidad en buenos resultados [2]. Merece la pena destacar que la fiabilidad y la validez no son características de los cuestionarios [3], 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 [4].

Aunque existen multitud de cuestionarios ya validados y aceptados por la comunidad científica en distintos ámbitos como la medicina, psicología, etc. en el caso de la educación son bastantes más escasos. Este hecho ha motivado que nos planteásemos describir el proceso de validación del cuestionario CoMaR (**Compresión y Manejo de la Recursividad**), para la identificación de las dificultades de los alumnos durante el proceso de aprendizaje de la recursividad, presentado en una edición anterior de las JENUI [5].

Por tanto, 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.

El artículo se estructura de la siguiente manera: en la sección 2 describimos cómo se mide la calidad de un cuestionario y el proceso recomendado para llevarlo a cabo; en la sección 3 detallamos cómo hemos llevado a cabo el análisis de la calidad del cuestionario CoMaR; en la sección 4 indicamos los principales resultados conseguidos para acabar, en la sección 5, con las principales conclusiones obtenidas.

## 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<sup>2</sup> o factores en las que se estructuren o agrupen 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 y/o imperfección asociado al mismo. Por ello, para el estudio de su calidad resulta imprescindible analizar su *validez* y su *fiabilidad*. Es de 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 se recomienda 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 cualificadas en el tema [6]. No obstante, en el caso en el que no exista acuerdo entre los jueces, se puede calcular el *índice de validez de contenido* [7], 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 variables<sup>3</sup> definen una estructura dimensional en el cuestionario que se mantenga invariante y pueda servir de base para la interpretación de los resultados en distintas poblaciones [3]. Previamente, y para comprobar la pertinencia de este

<sup>2</sup> Las *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

<sup>3</sup> El término *variables* es una forma abreviada de referirnos a las variables asociadas a las preguntas o ítems contenidas en el cuestionario

tipo de análisis, se recomienda calcular la *medida de adecuación muestral KMO* [8], 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 [9]. Además, es interesante confirmar el resultado anterior mediante la realización de la *prueba de esfericidad de Barlett* [10], 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.

Una vez confirmada la conveniencia de dicho análisis, éste suele llevarse a cabo con ayuda de los modelos de *análisis factorial exploratorio (AFE)* o de *análisis factorial confirmatorio (AFC)* [11]. 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. 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 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 cualquier caso, 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 [11]. En este caso, para evaluar la bondad del modelo se puede utilizar la prueba  $\chi^2$ ,

aunque por su sensibilidad al tamaño de la muestra se debería complementar con otros índices [12].

Por último, para facilitar la interpretación del modelo factorial obtenido y obtener una solución más acorde con el modelo previo, se puede realizar una *rotación* de la solución [11]. 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. En general, en el caso de los cuestionarios docentes, se prefiere la oblicua dada la correlación que suele existir entre sus dimensiones [13].

## 2.2. Fiabilidad

La evaluación de la *fiabilidad* de un cuestionario conlleva, entre otros, la realización de los siguientes análisis:

- **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* [14], que está basado en la correlación interelementos promedio y asume que los ítems (medidos en escala tipo Likert) miden un mismo constructo y que están altamente correlacionados [15]. Los valores de este coeficiente oscilan entre 0 y 1, considerando como criterio general un coeficiente *aceptable* cuando su valor es igual o superior a 0.70 [16] puesto que este estadístico no va acompañado de ningún *p-valor* (contraste de hipótesis). Además, se aconseja evaluar el valor del coeficiente *alfa* al eliminar del cuestionario cada uno de los ítems que lo componen, ya que se puede prescindir de aquellos en los que 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 *t de Student*, 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 [9].

### 3. Calidad del cuestionario CoMaR

#### 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 2º curso *Metodología de la Programación*, que se imparte en el 2º 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 [17], 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 [11]. 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 [11,13].

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.

|   |
|---|
| <b>Conocimiento Previo</b>  |
| P1. Conozco los <i>principios teóricos de la recursividad</i>   |
| P2. Comprendo los <i>principios teóricos de la recursividad</i>                                       |
| P3. Conozco el <i>funcionamiento de la recursividad</i>   |
| P4. Comprendo el <i>funcionamiento de la recursividad</i>   |
| P5. Conozco el principio de <i>inducción matemática</i>   |
| P6. Comprendo el principio de <i>inducción matemática</i>   |
| <b>Actitud</b>  |
| P7. Me gusta <i>programar</i>   |
| P8. En general, <i>me gusta</i> la recursividad   |
| <b>Utilidad percibida</b>   |
| P9. Considero <i>útil</i> la recursividad para la solución de problemas complejos                     |
| P10. <i>Valoro la utilidad</i> de la recursividad en programación                                     |
| <b>Dificultades Diseño</b>  |
| P11. Me cuesta trabajo <i>pensar de forma declarativa</i>   |
| P12. Me cuesta trabajo <i>diseñar la solución recursiva</i> de un problema                            |
| P13. Me cuesta trabajo <i>definir</i> el caso base  |
| P14. Me cuesta trabajo " <i>creerme</i> " el caso n-1   |
| P15. Me cuesta trabajo <i>definir</i> el caso general   |
| <b>Dificultades Análisis</b>  |
| P16. Me cuesta trabajo <i>analizar (tratar de entender)</i> el funcionamiento de programas recursivos |
| P17. Me cuesta trabajo <i>entender</i> el caso base   |
| P18. Me cuesta trabajo <i>entender</i> el caso n-1  |
| P19. Me cuesta trabajo <i>entender</i> el caso general  |
| <b>Dificultades Ejecución</b>   |
| 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 <i>varias llamadas recursivas</i>                     |
| P27. Seguimiento de la <i>traza</i> de un algoritmo recursivo   |

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

#### 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 pasó 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.

| 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        |
| <b>6</b> | <b>1.470</b> | <b>5.444</b>     | <b>63.926</b> |
| 7        | 1.245        | 4.613            | 68.538        |
| 8        | 1.046        | 3.873            | 72.411        |
| 9        | .916         | 3.393            | 75.804        |
| 10       | .819         | 3.034            | 78.839        |
| 11       | .723         | 2.676            | 81.515        |
| 12       | .629         | 2.330            | 83.845        |
| 13       | .600         | 2.223            | 86.068        |
| 14       | .543         | 2.011            | 88.079        |
| 15       | .485         | 1.795            | 89.874        |
| 16       | .445         | 1.648            | 91.522        |
| 17       | .380         | 1.406            | 92.929        |
| 18       | .356         | 1.317            | 94.246        |
| 19       | .290         | 1.074            | 95.319        |
| 20       | .250         | .926             | 96.245        |
| 21       | .202         | .747             | 96.992        |
| 22       | .190         | .704             | 97.697        |
| 23       | .185         | .687             | 98.383        |
| 24       | .161         | .596             | 98.979        |
| 25       | .114         | .424             | 99.403        |
| 26       | .089         | .329             | 99.732        |
| 27       | .072         | .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.

En cuanto al análisis factorial del cuestionario CoMaR, su objetivo es doble: 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. Antes de ello, calculamos la medida de adecuación muestral *KMO* y obtuvimos un valor de 0.656 (>0.5). Además, el valor del nivel de significación de la *prueba de esfericidad de Bartlett* es 0. Ambos datos sirven para confirmar que merece la pena realizar el análisis factorial. Hemos realizado un *análisis factorial confirmatorio (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 >1), 2 más de los que hemos definido inicialmente, por lo que decidimos analizar la fiabilidad del cuestionario por si éste sugiere la eliminación de algún ítem y, por ende, de algún factor.

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

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

### 3.3. Análisis de 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 aquéllos cuyo *índice de homogeneidad* es <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 tres ítems, P2, P6 y P9, que 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).

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

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

El índice de fiabilidad del cuestionario reducido a los 14 ítems restantes es de 0.825, un valor que se considera bueno [16]. Además, los índices de homogeneidad de todos los ítems son mayores o iguales a 0.2, enumerados en el Cuadro 5. Por tanto, podemos considerar que el cuestionario tiene una buena consistencia interna.

### 3.4. Análisis factorial exploratorio del cuestionario reducido

Finalizado el proceso de simplificación del cuestionario mediante los sucesivos análisis de fiabilidad, el siguiente paso es el de determinar el número óptimo de factores o dimensiones mediante un *análisis factorial exploratorio*, pues al haber sufrido el cuestionario una reducción de ítems tan drástica (prácticamente se han eliminado la mitad de los ítems) el modelo dimensional original ha dejado de tener sentido. El valor de la medida de adecuación muestral *KMO* ahora es de 0.741 (> 0.5), y mayor que para el 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.

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

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 que explican el 56% de la varianza total. Además, hemos seleccionado un método de rotación oblicua (Oblimin normal). Con estos datos, el índice de bondad del ajuste o *p-valor*, calculado mediante la prueba de  $\chi^2$ , es de 0.435. Como es mayor que 0.01, aceptamos la hipótesis nula con un nivel de confianza del 99%. Por tanto, 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. Las comunalidades son todas mayores que 0 aunque las más altas son las de los ítems P18, P21, P22 y P26, estando el valor de este último muy cercano a 1.

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

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

#### 4. Discusión de los resultados

El cuestionario ofrece 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. Este resultado confirma los resultados obtenidos en estudios previos [5] en los que no se encontró relación significativa entre dichos factores y los demás.

Por otra parte, la eliminación de las preguntas P13 y P17, 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 [5] 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 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*.

| Dificultades Conceptuales         |  |
|-----------------------------------|--|
| PR1.                              | Me cuesta trabajo <i>pensar de forma declarativa</i>   |
| PR2.                              | Me cuesta trabajo <i>diseñar la solución recursiva</i> de un problema                            |
| PR3.                              | Me cuesta trabajo " <i>creerme</i> " el caso n-1   |
| PR4.                              | Me cuesta trabajo <i>definir</i> el caso general   |
| PR5.                              | Me cuesta trabajo <i>analizar (tratar de entender)</i> el funcionamiento de programas recursivos |
| PR6.                              | Me cuesta trabajo <i>entender el caso n-1</i>  |
| PR7.                              | Me cuesta trabajo <i>entender el caso general</i>  |
| 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 <i>varias llamadas recursivas</i>                     |
| PR14.                             | Seguimiento de la <i>traza</i> 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.

En consecuencia, el cuestionario CoMaR\* de 14 preguntas, descrito en el Cuadro 7, ofrece mayor calidad 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 son más válidos y más fiables 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 el conocimiento estadístico del investigador para garantizar la calidad de los análisis y la validez de los

resultados obtenidos. Por tanto, creemos que este trabajo aporta a la comunidad docente 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.

La segunda limitación está relacionada con la muestra del estudio, porque aunque su tamaño es aceptable [18], no es representativa de todos los estudiantes de programación, por lo que las conclusiones obtenidas son solo significativas para la muestra del estudio.

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 [11].

## Referencias

- [1] C. Ruiz Bolívar, "Validez," UPEL / PIDE, 2000. [http://investigacion.upeu.edu.pe/images/7/74/V\\_alidez.pdf](http://investigacion.upeu.edu.pe/images/7/74/V_alidez.pdf)
- [2] I. Lucero and S. Meza, "Validación de instrumentos para medir conocimientos," Departamento de Física - Facultad de CC. Exactas y Naturales y Agrimensura, UNNE, Argentina, 2002. [www1.unne.edu.ar/cyt/2002/09-Educacion/D-027.pdf](http://www1.unne.edu.ar/cyt/2002/09-Educacion/D-027.pdf)
- [3] G. Delgado, A. Prieto, "Fiabilidad y Validez," *Papeles del Psicólogo*, vol. 31, no. 1, pp. 67-74, 2010.
- [4] 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.
- [5] C. Lacave, A.I. Molina, and J. Giralt, "Identificando algunas causas del fracaso en el aprendizaje de la recursividad. Análisis experimental en las asignaturas de programación.," *Actas de las XIX JENUI, Castellón*, pp. 225-232, 2013.
- [6] Cuervo-Martínez, A. Escobar-Pérez, J, "Validez de contenido y juicio de expertos: una aproximación a su utilización," *Avances en Medición*, vol. 6, pp. 27-36, 2008.
- [7] C.H. Lawshe, "A quantitative approach to content validity," *Personnel Psychology*, vol. 28, pp. 563-575, 1975.
- [8] H.F. Kaiser, "An index of factorial simplicity," *Psychometrika*, vol. 39, no. 1, pp. 31-36, 1974.
- [9] M.I Barbero García, E. Vila Abad, and J.C. Suárez Falcón, *Psicometría*. Nardrud: UNED, 2006.
- [10] Bartlett, "Tests of significance in factor analysis," *British Journal of Statistical Psychology*, vol. 3, no. 2, pp. 77-85, 1950.
- [11] P.J Ferrando and C Anguiano-Carrasco, "El análisis factorial como técnica de investigación en Psicología," *Papeles del Psicólogo*, vol. 31, no. 1, pp. 18-33, 2010.
- [12] M. Fernández et al., "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 on Educational Psychology*, vol. 8, no. 3, pp. 1229-1252, 2010.
- [13] D Frías-Navarro and M Pascual Soler, "Prácticas del análisis factorial exploratorio (AFE) en la investigación sobre conducta del consumidor y marketing," *Suma Psicológica*, vol. 19, no. 1, pp. 47-58, 2012.
- [14] L.J. Cronbach, "Coefficient alpha and internal structure of test," *Psychometrika*, vol. 16, pp. 297-334, 1951.
- [15] S Welch and J.C. Comer, *Quantitative methods for public administration: techniques and applications.*: Dorsey Press, 1988.
- [16] D George and P. Mallery, *SPSS for Windows step by step: A simple guide and reference. 11.0 Update*, 1104th ed. Boston: Allyn & Bacon, 2003.
- [17] R Likert, "A technique for the measurement of attitudes," *Archives of Psychology*, vol. 22, no. 140, pp. 44-53, 1932.
- [18] Pedro Morales Vallejo, "Guía para construir cuestionarios y escalas de actitudes," Universidad Pontificia Comillas, Madrid, 2011. <http://web.upcomillas.es/personal/peter/otrosdocumentos/Guiaparaconstruיריםcalasdeactitudes.pdf>