

MEJORANDO LA VALIDEZ Y FIABILIDAD DE INSTRUMENTOS DE EVALUACIÓN MEDIANTE EL ANÁLISIS ESTADÍSTICO IMPLICATIVO

Silvia MAYÉN¹ y Carmen DÍAZ²

TITRE

Amélioration de la validité et de la fiabilité d'outils d'évaluation par l'Analyse Statistique Implicative

TITLE

Improving the validity and reliability of assessment tools by Implicative Analysis

RESUMEN

El propósito de este trabajo es mostrar la utilidad del Análisis Implicativo como método estadístico que refuerza la validez y fiabilidad de instrumentos de evaluación obtenida con otros métodos. En este trabajo presentamos el análisis implicativo a las respuestas de un cuestionario que evalúa la comprensión de medidas de tendencia central en estudiantes mexicanos de educación secundaria y bachillerato, y que por lo tanto, puede ser generalizable a estudiantes de otros contextos y niveles educativos.

Palabras clave: análisis implicativo, validez, fiabilidad, medidas de tendencia central

RÉSUMÉ

Le but de cet article est de montrer l'utilité de l'analyse implicative comme méthode statistique renforçant la validité et la fiabilité des instruments d'évaluation obtenus avec d'autres méthodes. Les réponses d'élèves mexicains de collège et lycée à un questionnaire pour évaluer leur compréhension des mesures de tendance centrale ont été soumises à une analyse implicative, et nous voyons une possibilité de généralisation à d'autres contextes ou d'autres niveaux scolaires.

Mots-clés : analyse implicative, validité, fiabilité, mesures de tendance centrale.

ABSTRACT

The purpose of this paper is to show the utility of implicative analysis as a statistical method that reinforces the validity and reliability of assessment tools obtained with other methods. A questionnaire to assess understanding the central tendency measures was applied to Mexican secondary and high school students, and their answers were submitted to Implicative Analysis; obtained results can be generalized to students from different contexts and educational levels.

Keywords: implicative analysis, validity, reliability, measures of central tendency.

¹ Centro de Investigación y de Estudios Avanzados, Avenida Instituto Politécnico Nacional 2508, Col. San Pedro Zacatenco, 07360, México, D. F., mayazuc@gmail.com

² Universidad de Huelva, Avenida Fuerzas Armadas s/n, 21071, Huelva, Es., Carmen.diaz@dpsi.uhu.es

1 Introducción

Este trabajo forma parte de un estudio más amplio acerca de un análisis cuantitativo de evaluación sobre la comprensión de las medidas de tendencia central en estudiantes mexicanos, que consiste en la aplicación de un cuestionario a diferentes grupos de estudiantes. Incluye la validación de dicho cuestionario mediante tres tipos de evidencia de validez: 1) validez de contenido, justificada mediante análisis teórico de los ítems; 2) validez discriminante, mediante análisis de diferencia de ejecución en los ítems en los grupos; y 3) validez de constructo, analizando la estructura de las respuestas mediante análisis clúster e implicativo. Así mismo, se lleva a cabo la aproximación a la fiabilidad mediante el coeficiente Alfa (Martínez Arias, 1995), coeficiente Theta (Barbero, 2003) y Teoría de la Generalizabilidad (Feldt y Brennan, 1991). Abarca también un estudio global de la dificultad de los ítems del cuestionario en la muestra y una comparación ítem a ítem de los resultados en los distintos grupos de estudiantes.

Así, en esta comunicación centraremos nuestra atención en el Análisis clúster y Análisis implicativo de dicho estudio, y para su mejor comprensión, se hace necesario mostrar los resultados de los análisis mencionados anteriormente, la descripción de las muestras de estudiantes y sus contextos, así como explicar las características del cuestionario y la metodología de su aplicación.

2 Estudio cuantitativo de evaluación

El origen de este estudio es la investigación previa de Cobo (2003) sobre la comprensión de las medidas de tendencia central. La autora se interesó en comparar el nivel de conocimientos sobre dicho tema en estudiantes españoles al comienzo y al final de la Educación Secundaria Obligatoria, de edades comprendidas entre 12 y 16 años. Con este propósito, Cobo (2003) construyó un cuestionario compuesto por dieciséis ítems abiertos relacionados con las medidas de tendencia central, a partir de distintas investigaciones y considerando los contenidos curriculares españoles.

En nuestro caso, incluimos alumnos mexicanos de 13 a 19 años y nos interesamos específicamente por los estudiantes que finalizan la Educación Secundaria y también el Bachillerato, por ser dos niveles importantes en el sistema educativo mexicano, en cuanto que los alumnos se preparan para un cambio de nivel, es decir, para pasar de Secundaria a Bachillerato o de Bachillerato a la Universidad o Formación Profesional en cada caso. El objetivo principal es analizar si las dificultades citadas por la autora se resuelven, se repiten o surgen otras nuevas en los estudiantes de cursos superiores, así como evaluar sus conocimientos acerca de las medidas de tendencia central. La elección de este instrumento se debe a que los ítems presentan enunciados de situaciones comprensibles y familiares para nuestros estudiantes, y también recoge los contenidos de los programas de estudios mexicanos de ambos niveles sobre medidas de centralización.

Queremos resaltar que en nuestro estudio entendemos la *validación de un cuestionario* como un proceso por el cual se aportan evidencias que apoyen la interpretación propuesta de los datos recogidos mediante la prueba (Carmona, 2004) y

que está basada tanto en los procesos de respuesta, como en las consecuencias de la evaluación (AERA/ APA/ NCME, 1999). En nuestro caso, la consecuencia de la evaluación es el diagnóstico de los conocimientos de los estudiantes. Además, asumimos una visión unificada del concepto de *validez* propuesto por Messick (1998), es decir, asumimos la validación como un proceso continuo, puesto que cualquier cuestionario puede mejorarse en el futuro.

2.1 Descripción de la muestra

La muestra está compuesta por 518 estudiantes mexicanos de entre 13 y 19 años de edad: 162 estudiantes de Educación Secundaria y 356 de Bachillerato. Los estudiantes de Secundaria tienen entre 14 y 15 años de edad y son del tercer curso de dicho nivel, por lo que podemos considerarlos equivalentes al tercer nivel de la Educación Secundaria Obligatoria en España. La otra parte de la muestra está compuesta por estudiantes del tercer año de Bachillerato, y la mayoría son de 17 y 18 años de edad.

Incluimos también, una segunda muestra de estudiantes de un estudio piloto (Mayén, 2006), que tenía como finalidad comprobar que el cuestionario de Cobo (2003) era adecuado para ser utilizado en el contexto mexicano, en cuanto a su dificultad y contextos de los ítems, y con estudiantes de mayor edad. Así también, si los estudiantes proporcionan respuestas suficientemente claras y completas para realizar nuestra investigación posterior. Esta muestra se compone de 125 estudiantes de bachillerato, y ha sido necesario considerarla para obtener una estimación más precisa de la fiabilidad, al contar con una muestra de mayor tamaño.

Cabe señalar que todos los grupos de estudiantes tenían conocimientos del tema de medidas de tendencia central.

2.2 El cuestionario

El *cuestionario* es un instrumento de medición que por medio de las preguntas planteadas se obtiene una estimación de conocimientos y capacidades de los sujetos a quienes se les aplica, que no son accesibles por simple observación o encuesta (Dane, 1990; Barbero, 2003).

El cuestionario que utilizamos para nuestra investigación está orientado a la evaluación del significado que estudiantes mexicanos asignan a las medidas de tendencia central. Fue construido por Cobo (2003) después de un análisis sistemático del contenido de las medidas de posición central en una amplia muestra de libros de texto españoles de secundaria, así como en el análisis de las directrices curriculares españolas. En nuestro caso, también hemos analizado el currículo mexicano de educación secundaria y de bachillerato. Así mismo, queremos informar sobre la comprensión de los siguientes tipos de elementos, que quedan establecidos en el cuestionario:

- Reconocimiento de los campos de problemas que se resuelven mediante promedios.
- Uso por parte de los alumnos de las diferentes definiciones de media, mediana y moda en la solución de los problemas propuestos o en la justificación de la solución.

- Comprensión de propiedades básicas, tanto numéricas, como algebraicas y estadísticas; uso adecuado de dichas propiedades al responder a las preguntas planteadas.
- Reconocimiento del lenguaje matemático verbal, numérico y gráfico; uso apropiado de términos y lenguaje.
- Cálculo y procedimientos de resolución de problemas. Comprensión de los algoritmos de cálculo frente a su aplicación automática.
- Argumentaciones de los alumnos para apoyar sus respuestas y observar hasta qué punto son completas y consistentes.

Para su elaboración, Cobo (2003) siguió las recomendaciones de Osterlind (1989), Thorndike (1989) y Linn (1988), en cuanto a redacción de los ítems, formato, claridad y orden de presentación. También tuvo en cuenta el *Tipo de administración*: en forma grupal, y *Limitaciones temporales en la administración*: se hizo un balance entre el número suficiente de ítems (y su dificultad) y el tiempo disponible para contestarlo.

El cuestionario incluye un total de 16 ítems, algunos de ellos divididos en subítems (en total 27 subítems), por lo tanto, se puede obtener una puntuación total en el cuestionario que varía de 0 a 27, y todos los ítems son de respuesta abierta para poder recoger con detalle los razonamientos de los estudiantes.

Al tratar de evaluar la comprensión, se tuvo en cuenta que es un constructo inobservable (León y Montero, 2002), por lo que sus características deben ser inferidas de las respuestas de los alumnos. La comprensión de los estudiantes sobre un cierto objeto matemático (en este caso las medidas de tendencia central) es inobservable. Pero las prácticas que realizan al resolver problemas, y en particular los problemas presentados como ítems en un cuestionario, sí que son observables, siempre que la recogida de datos sea completa y fiable (Godino, 1996).

El cuestionario tiene, por tanto, como principal objetivo, recoger datos sobre las prácticas matemáticas que realizan los estudiantes al resolver problemas relacionados con las medidas de posición central para aproximarnos a dicha comprensión. De las respuestas escritas trataremos de inferir el uso (correcto o incorrecto) que los estudiantes de la muestra hacen de los diversos objetos matemáticos: lenguaje, definiciones, propiedades, procedimientos y argumentos.

Las interpretaciones realizadas a partir de las respuestas harán referencia a lo que los sujetos hacen o son capaces de hacer, y a sus conocimientos y errores sobre el mismo (test referido a criterio). Se trata entonces de un cuestionario de potencia, ya que el tiempo, aunque controlado, no determinaría el resultado, sino que las diferencias en la puntuación serían debidas a la calidad de su ejecución y conocimiento (Sax, 1989; Martínez Arias, 1995). El cuestionario propuesto se incluye en la sección de Anexos.

2.3 Resultados del Estudio Cuantitativo de Evaluación

En esta sección se describe el desarrollo del estudio cuantitativo, que toma en cuenta las respuestas al cuestionario como correctas o incorrectas, considerando las respuestas en blanco como incorrectas. A partir de esta primera codificación analizamos la *dificultad comparada de los ítems* y el rendimiento total de los estudiantes en la prueba, así como las características psicométricas del cuestionario para completar su validación. Los resultados los presentamos a continuación.

2.3.1 Validez de contenido del cuestionario

La validez de contenido es el grado en que el instrumento de evaluación refleja el dominio que nos interesa en forma satisfactoria (Carmines y Zeller, 1979). Se trata de ver la adecuación de los ítems de un test como muestra de un universo más amplio de ítems representativos del contenido (Martínez Arias, 1995). Siguiendo a esta autora la validación del contenido se ha hecho mediante el examen sistemático del contenido del test para probar su representatividad y relevancia. Con ello comprobamos que los ítems de nuestro test son relevantes para el uso que se dará a las puntuaciones y representativos del contenido que se quiere evaluar, representando sus características esenciales.

2.3.2 Estudio global de resultados

El estudio cuantitativo tiene como finalidad presentar los resultados obtenidos de las respuestas de los estudiantes al cuestionario administrado. Se lleva a cabo para la totalidad de la muestra y comparando los grupos que la componen, es decir, entre estudiantes de Secundaria y de Bachillerato. El estudio global de resultados se organiza en distintos apartados: Análisis comparado de la dificultad de los ítems; Cálculo de los intervalos de confianza y credibilidad para estos índices, los que completamos con el uso de métodos bayesianos para mejorar la estimación de los índices de dificultad siguiendo el modelo implementado por Díaz (2007) en la construcción de un cuestionario de evaluación de la comprensión de la probabilidad condicional; Análisis de la puntuación total y discusión de los resultados sobre la misma; y Estudio de fiabilidad y generalizabilidad del cuestionario.

2.3.3 Dificultad comparada de ítems

Para realizar el análisis de dificultad de los ítems del cuestionario, tomamos las respuestas de los alumnos, y aunque el cuestionario contiene solamente ítems abiertos, en esta parte de la investigación analizamos primeramente si el alumno es capaz o no de dar las respuestas correctas esperadas, las cuales categorizamos únicamente como correctas o incorrectas.

En la Tabla 1 (sección 2.3.4), se presentan para el total de la muestra los índices de dificultad de cada ítem, entendiéndolo en la acepción de Muñiz (1994), como la proporción de sujetos que lo aciertan entre todos los que trataron de resolverlo. Cuanto mayor es este valor, significa que el ítem es más fácil para los alumnos y ha sido respondido correctamente por una mayor proporción de ellos. Este índice fluctúa entre 0.24 en el ítem 2.3 (cálculo de la media de una suma de variables), 0.26 en el ítem 10.1 (cálculo de la media, mediana y moda de un conjunto de datos agrupados en intervalos y presentados en una tabla de frecuencias absolutas), y 0.85 en el ítem 8 (estimación de una cantidad desconocida a partir de diversas mediciones en presencia de errores).

La mayor parte de los ítems tiene una dificultad moderada, en concreto, 24 de los 27 subítems tienen un índice de dificultad comprendido entre 0.3 y 0.7. Con ello conseguiremos más discriminación entre los estudiantes, y en definitiva mejores resultados en la evaluación. El análisis de estas tablas y gráficos muestra que algunos ítems son difíciles de resolver para el global de alumnos.

2.3.4 Estimaciones bayesianas

Utilizamos en nuestra investigación los métodos bayesianos como complemento de los métodos clásicos de inferencia, por algunas razones como la interpretación más intuitiva de los resultados proporcionados (Díaz, 2007), y sobre todo, la posibilidad de tener en cuenta la información previa que se posea sobre la población en estudio. En nuestro caso no sería lógico proceder como si no se dispusiese de información anterior sobre la dificultad de los ítems del cuestionario o como si todos ellos tuviesen la misma probabilidad de ser resueltos por nuestros estudiantes. El estudio previo de Cobo (2003) ofrece un análisis detallado de la dificultad de estos ítems en dos muestras de estudiantes de Educación Secundaria en España, siendo el segundo grupo muy semejante en edad a nuestro grupo de estudiantes de Secundaria y habiendo seguido un número similar de cursos de matemáticas previos. Además, el contenido estudiado sobre las medidas de tendencia central también es muy similar en ambos países.

TABLA 1- *Índice de dificultad, Intervalos de Confianza y Credibilidad del 95%*

| Ítem | Estimación | | Intervalo de credibilidad | | |
|-------|-------------------|------------------------|---------------------------|----------------|-------------|
| | Clásica | Intervalo de confianza | | No informativo | |
| | Índice dificultad | L. inferior | L. superior | L. inferior | L. superior |
| I1_1 | 0.71 | 0.671 | 0.749 | 0.670 | 0.748 |
| I1_2 | 0.65 | 0.610 | 0.692 | 0.608 | 0.690 |
| I2_1 | 0.30 | 0.260 | 0.339 | 0.261 | 0.340 |
| I2_2 | 0.37 | 0.329 | 0.412 | 0.330 | 0.413 |
| I2_3 | 0.24 | 0.203 | 0.276 | 0.204 | 0.278 |
| I3 | 0.65 | 0.610 | 0.692 | 0.608 | 0.690 |
| I4 | 0.76 | 0.724 | 0.797 | 0.722 | 0.795 |
| I5_1 | 0.68 | 0.639 | 0.720 | 0.639 | 0.719 |
| I5_2 | 0.60 | 0.558 | 0.643 | 0.558 | 0.642 |
| I5_3 | 0.31 | 0.271 | 0.351 | 0.272 | 0.351 |
| I6_1 | 0.37 | 0.329 | 0.412 | 0.330 | 0.413 |
| I6_2 | 0.36 | 0.318 | 0.400 | 0.319 | 0.401 |
| I7_1 | 0.78 | 0.744 | 0.816 | 0.743 | 0.814 |
| I7_2 | 0.76 | 0.724 | 0.797 | 0.722 | 0.795 |
| I7_3 | 0.68 | 0.639 | 0.720 | 0.639 | 0.719 |
| I8 | 0.85 | 0.819 | 0.880 | 0.817 | 0.879 |
| I9_1 | 0.63 | 0.588 | 0.671 | 0.588 | 0.671 |
| I9_2 | 0.32 | 0.280 | 0.361 | 0.281 | 0.361 |
| I10_1 | 0.26 | 0.223 | 0.298 | 0.224 | 0.299 |
| I10_2 | 0.30 | 0.260 | 0.339 | 0.261 | 0.340 |
| I10_3 | 0.59 | 0.548 | 0.633 | 0.548 | 0.632 |
| I11 | 0.34 | 0.299 | 0.381 | 0.300 | 0.381 |
| I12 | 0.27 | 0.671 | 0.749 | 0.670 | 0.748 |
| I13 | 0.37 | 0.610 | 0.692 | 0.608 | 0.690 |
| I14 | 0.34 | 0.260 | 0.339 | 0.261 | 0.340 |
| I15 | 0.57 | 0.329 | 0.412 | 0.330 | 0.413 |
| I16 | 0.35 | 0.203 | 0.276 | 0.204 | 0.278 |

En consecuencia, en esta sección completamos nuestro estudio mediante la estimación clásica y bayesiana de los índices de dificultad de los ítems que conforman el cuestionario y usando la información previa obtenida en el estudio de Cobo (2003).

De esta forma mejoramos nuestras estimaciones y ofrecemos una interpretación más natural de los intervalos en torno a estas estimaciones (intervalos de credibilidad). Así, en la Tabla 1 aparecen los *intervalos de confianza* (estimación clásica) y los intervalos de credibilidad (estimación bayesiana) para los índices de dificultad de cada uno de los ítems. El primero de ellos ha sido calculado con la fórmula ordinaria de intervalos de confianza de una proporción y tiene una interpretación frecuencial, es decir, en cada 100 muestras tomadas de la misma población, 95% de ellas contendrían la proporción verdadera, aunque no podemos saber si se contiene o no en nuestra muestra particular.

El *intervalo de credibilidad*, indica el intervalo de valores en que esperamos que la proporción verdadera esté incluida, es decir, nos da una probabilidad epistémica, que se refiere a la muestra particular. Para analizar las diferencias que introducen las estimaciones bayesianas, consideramos, una distribución inicial uniforme, es decir, supusimos equiprobables todos los valores de la proporción a priori en los diferentes ítems. Por ello, los intervalos de confianza y credibilidad son muy similares.

En las Tablas 2^a y 2B presentamos las estimaciones bayesianas, en este caso con la distribución inicial informativa (estudio de Cobo, 2003). Observamos que ahora los intervalos de credibilidad obtenidos son más precisos. El valor estimado se corrige con la información previa, aunque, como nuestra muestra es mayor, en caso de diferencia entre las dos estimaciones, el valor final se aproxima más al de nuestro estudio.

TABLA 2A - *Estimación bayesiana de índices de dificultad con distribución informativa*

| | Índice observado (n=518) | Proporción en Cobo (2003) 4°ESO, n=144 | Estimación bayesiana del índice de dificultad | | |
|------|-----------------------------|---|---|-------------|-------------|
| | | | Valor medio | L. inferior | L. superior |
| I1_1 | 0.71 | 0.69 | 0.706 | 0.670 | 0.740 |
| I1_2 | 0.65 | 0.37 | 0.589 | 0.551 | 0.626 |
| I2_1 | 0.30 | 0.34 | 0.308 | 0.274 | 0.344 |
| I2_2 | 0.37 | 0.38 | 0.373 | 0.336 | 0.410 |
| I2_3 | 0.24 | 0.33 | 0.260 | 0.227 | 0.294 |
| I3 | 0.65 | 0.49 | 0.616 | 0.578 | 0.652 |
| I4 | 0.76 | 0.66 | 0.738 | 0.704 | 0.771 |
| I5_1 | 0.68 | 0.38 | 0.615 | 0.577 | 0.651 |
| I5_2 | 0.60 | 0.32 | 0.539 | 0.501 | 0.577 |
| I5_3 | 0.31 | 0.33 | 0.315 | 0.280 | 0.351 |
| I6_1 | 0.37 | 0.13 | 0.318 | 0.284 | 0.354 |
| I6_2 | 0.36 | 0.04 | 0.290 | 0.256 | 0.325 |
| I7_1 | 0.78 | 0.67 | 0.756 | 0.722 | 0.788 |
| I7_2 | 0.76 | 0.68 | 0.743 | 0.709 | 0.775 |
| I7_3 | 0.68 | 0.61 | 0.665 | 0.628 | 0.700 |
| I8 | 0.85 | 0.67 | 0.811 | 0.780 | 0.840 |
| I9_1 | 0.63 | 0.67 | 0.638 | 0.601 | 0.674 |
| I9_2 | 0.32 | 0.26 | 0.307 | 0.272 | 0.342 |

TABLA 2B - *Estimación bayesiana de índices de dificultad con distribución informativa*

| | Índice observado (n=518) | Proporción en Cobo (2003) 4ºESO, n=144 | Estimación bayesiana del índice de dificultad | | |
|-------|-----------------------------|---|---|-------------|-------------|
| | | | Valor medio | L. inferior | L. superior |
| I10_1 | 0.26 | - | 0.261 | 0.224 | 0.299 |
| I10_2 | 0.30 | - | 0.300 | 0.261 | 0.340 |
| I10_3 | 0.59 | - | 0.590 | 0.548 | 0.632 |
| I11 | 0.34 | 0.26 | 0.322 | 0.287 | 0.358 |
| I12 | 0.27 | 0.15 | 0.245 | 0.213 | 0.278 |
| I13 | 0.37 | 0.34 | 0.364 | 0.328 | 0.401 |
| I14 | 0.34 | 0.27 | 0.325 | 0.290 | 0.361 |
| I15 | 0.57 | 0.45 | 0.544 | 0.506 | 0.582 |
| I16 | 0.35 | 0.49 | 0.381 | 0.344 | 0.418 |

Con estos resultados finalizamos el estudio bayesiano de los índices de dificultad del cuestionario, complementando de este modo el estudio de las características del cuestionario.

2.3.5 Análisis de la puntuación total en el cuestionario

Para analizar el número de respuestas correctas que cada estudiante obtuvo, asignamos el valor 1 a cada respuesta correcta, por lo que al sumarlas podríamos encontrar una repuntuación total entre 0 y 27 puntos. Observamos en la Tabla 3 que el número de respuestas correctas osciló entre 5 y 23, por lo que ningún estudiante llega a dar todas las respuestas correctas, aunque todos tienen al menos cinco correctas. El número medio de respuestas correctas es de 13.42 sobre 27, lo que nos da resultados aceptables acercándose a lo esperado, pues el punto medio sería 13.5. La mediana (Mdn=13) indica que la mitad de los estudiantes responden al menos la mitad de los ítems del cuestionario.

TABLA 3 - *Estadísticos descriptivos de la puntuación total*

| | Estadístico | Error típ. |
|---|------------------------------------|----------------|
| Media | 13.42 | 0.168 |
| Intervalo de confianza para la media al 95% | Límite inferior Límite superior | 13.08 13.75 |
| Mediana | 13.00 | |
| Varianza | 14.622 | |
| Desv. típ. | 3.824 | |
| Mínimo | 5 | |
| Máximo | 23 | |
| Amplitud intercuartil | 5 | |
| Asimetría | 0.074 | 0.107 |
| Curtosis | -0.583 | 0.214 |

Todo ello nos asegura unas buenas características del cuestionario, ya que nos permite discriminar una amplia gama de conocimientos, que van desde muy bajos hasta conocimientos altos, encontrándose la mayoría de los estudiantes alrededor del centro, es decir, contestando correctamente a la mitad de las preguntas. El conjunto central de alumnos (50% central), responde entre 10 y 16 preguntas correctamente de un total posible de 27 puntos; no hay sujetos atípicos, lo que es otro indicador de las buenas características del cuestionario. En todo caso hay una gran variabilidad en el número de respuestas correctas, lo que indica una comprensión desigual de las medidas de centralización por estos estudiantes. También supone un buen poder discriminador del cuestionario.

2.3.6 Fiabilidad y generalizabilidad

Análisis de la fiabilidad del cuestionario en nuestra muestra.

Cobo (2003) informó sobre este punto como la extensión por la cual un experimento, test u otro procedimiento de medida produce los mismos resultados en ensayos repetidos. La medida siempre produce un cierto error aleatorio, pero dos medidas del mismo fenómeno sobre un mismo individuo suelen ser consistentes. Sin embargo, la fiabilidad varía al cambiar la población objeto de estudio. Siguiendo a Thorndike (1989), evaluamos conceptos abstractos, en nuestro caso, relacionar el significado que los alumnos de la muestra asignan a las medidas de posición central con sus respuestas a los ítems del cuestionario, que son indicadores empíricos. Para permitir este proceso de medida, un indicador debe ser fiable. La *fiabilidad* es esta tendencia a la consistencia o precisión del instrumento en la población medida (Bisquerra, 1989). Para el cálculo de fiabilidad y generalizabilidad usamos todos los datos recogidos, que describimos en la sección 2.1, es decir, que en total configuran un grupo de 642 estudiantes con el fin de obtener una estimación más precisa de la fiabilidad. Para estimar la fiabilidad de una escala realizamos los siguientes pasos:

- En primer lugar, consideramos el método de *consistencia interna*, que está basado sólo en la aplicación del cuestionario (Díaz, Batanero y Cobo, 2003). Su cálculo se basa en el análisis relativo de la varianza de la puntuación total del cuestionario y de las varianzas de los ítems particulares; el coeficiente que lo mide es el Alfa de Cronbach (Carmines y Zeller, 1979). Obtuvimos un valor Alfa = 0.662 para el coeficiente de Cronbach, que se considera como un valor adecuado aunque no excesivamente elevado debido a que, como se probó en el estudio de Cobo (2003), el cuestionario evalúa un constructo que no es unidimensional. Morales (1988, pag. 249) indica que no hay regla fija para cuál debe ser el valor mínimo del coeficiente y que diversos autores han dado distintas reglas, como Santisteban (1990), que indica como límite general, 0.50. En todo caso, puesto que el cuestionario incluye una gama amplia de conceptos, no es de esperar una fiabilidad excesivamente alta.
- Seguidamente, calculamos un coeficiente de fiabilidad basado en el *análisis factorial*, siguiendo las consideraciones indicadas en Barbero (2003). Puesto que asumimos que el cuestionario evalúa un constructo multidimensional, la fiabilidad se calcula más exactamente con este coeficiente, que representa la contribución del primer factor del análisis factorial al total de la fiabilidad del cuestionario. Primeramente, en la Tabla 4 hacemos la prueba KMO y la prueba de Bartlett para comprobar que podemos aplicar el análisis factorial. Se presentan los resultados del análisis factorial con los 11 factores de autovalor mayor que uno. Esto confirma

nuestra hipótesis de multidimensionalidad del constructo. Al mismo tiempo, puesto que el primer factor tiene mayor varianza que el resto (los cuales explican casi la misma varianza cada uno) sugiere la existencia del constructo general “comprensión de las medidas de tendencia central” que se complementará después con el resto de factores. Todo ello apoya la validez de constructo del cuestionario.

TABLA 4 - *KMO y prueba de Bartlett*

| | |
|--|--------------|
| Medida de adecuación muestral de Kaiser-Meyer-Olkin. | 0.648 |
| Prueba de esfericidad de Bartlett | Chi-cuadrado |
| | 4844.404 |
| | Gl |
| | 351 |
| | Sig. |
| | 0.000 |

A partir de los resultados del análisis factorial se calculó el coeficiente Theta de Carmines³, siguiendo la fórmula (1):

$$\theta = \frac{n}{n-1} \left(1 - \frac{1}{\lambda_1} \right) = 0.726 \quad (1)$$

donde n es el número de ítems y λ el primer autovalor en el análisis factorial. Coincide con el coeficiente α calculado con las puntuaciones factoriales derivadas del primer factor común y sintetiza la información aportada por el primer factor (Morales, 1988).

En nuestro estudio el coeficiente alfa presenta un valor bastante alto, lo que junto con el hecho de que el primer factor explicó mucho mayor porcentaje de varianza que los siguientes, y que la mayoría de los ítems tenían pesos apreciables en el mismo, antes de la rotación, indica una cierta unidimensionalidad del cuestionario aunque pequeña, ya que el porcentaje de varianza explicado por el primer factor fue sólo el 13%.

Coefficientes de generalizabilidad.

La teoría de la generalizabilidad extiende la teoría clásica de la medición, según Feldt y Brennan (1991), y permite por medio del análisis de varianza, analizar diferentes fuentes de error en un proceso de medida. Para Santisteban (1990), el núcleo de esta teoría es considerar diferentes fuentes de error en las puntuaciones observadas, que pueden ser los mismos sujetos, las preguntas o las condiciones que se aplican. El coeficiente de generalizabilidad se define con el cociente (2), es decir, como el cociente entre la varianza verdadera en las puntuaciones de la prueba y la varianza observada, que es la suma de la varianza verdadera más la varianza debida al error aleatorio.

$$G = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_e^2} \quad (2)$$

En este trabajo hemos diferenciado dos fuentes para el error aleatorio y calculado la generalizabilidad de los mismos sujetos (inter-personas) y la generalizabilidad de los ítems (inter-elementos). Para realizar este cálculo, hemos considerado el modelo de estimación de Dunn y Clarck (1987) para el análisis de varianza de medidas repetidas, los siguientes componentes de la varianza (Tabla 5), tomando el ítem como factor.

³ Carmines y Zéller (1979) definen este coeficiente para cuestionarios no unidimensionales, como es el nuestro.

TABLA 5 - Análisis de varianza de medidas repetidas

| | | Suma de cuadrados | Gl | Media cuadrática | F | Sig |
|----------------|-----------------|----------------------|-------|---------------------|--------|-----|
| Inter-personas | | 374.074 | 642 | 0.583 | | |
| Intra-personas | Inter-elementos | 670.563 | 26 | 25.791 | 130.86 | 0.0 |
| | Residual(a) | 3289.733 | 16692 | 0.197 | | |
| | Total | 3960.296 | 16718 | 0.237 | | |
| Total | | 4334.370 | 17360 | 0.250 | | |

Media global = 0,48

$CM_S = 0.583$ que es un estimador de $b\sigma_S^2 + \sigma_R^2$, siendo b el número de ítems
 $CM_I = 25.791$ que es un estimador de $a\sigma_I^2 + \sigma_R^2$ siendo a el número de sujetos
 $CM_R = 0.197$ que es un estimador de σ_R^2

De donde, despejando obtenemos las siguientes estimaciones:

Varianza dentro de los sujetos $\sigma_s^2 = 0.0148$

Varianza dentro de los ítems $\sigma_i^2 = 0.0398$

Varianza residual $\sigma_e^2 = 0.197$

Sustituyendo ahora estos componentes de varianza en la fórmula (2) y teniendo en cuenta los tamaños de muestra (27 ítems y 642 alumnos), según si consideramos como fuente de variación los problemas o los alumnos, obtenemos las siguientes estimaciones:

$$\text{Generalizabilidad respecto a los ítems: } G_i = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_e^2/27} = 0.6616 \quad (3)$$

Obtenemos un valor próximo al del coeficiente Alfa (0.662), debido a que el coeficiente de generalizabilidad respecto a los ítems coincide con él (3), ya que se considera el número de alumnos fijo y la única fuente de variación se debe a la variabilidad entre ítems. Las pequeñas diferencias se deben a redondeos en los cálculos.

$$\text{Generalizabilidad respecto a los alumnos: } G_s = \frac{\sigma_i^2}{\sigma_i^2 + \sigma_e^2/642} = 0.9619 \quad (4)$$

Obtenemos un valor muy alto para la generalizabilidad a otros alumnos de la misma prueba (4), es decir, se indica una muy alta posibilidad de generalizar nuestros resultados a otros alumnos conservando el mismo cuestionario, por supuesto, bajo la hipótesis de que se conserven las características sociológicas y educativas.

3 Análisis Clúster y Análisis Implicativo

Para estudiar las interrelaciones entre objetivos de aprendizaje hemos llevado a cabo varios análisis multivariantes de las respuestas a los ítems de la prueba. Los resultados de estos análisis se presentan mediante análisis cluster, grafo implicativo y análisis implicativo jerárquico.

3.1 Análisis Clúster

Este análisis se llevó a cabo utilizando el software CHIC, Classification Hierarchical, Implicative et Cohesive (Couturier y Gras, 2005), que realiza un estudio de aglomeración jerárquica, tomando como medida de similaridad entre ítems el índice de Lerman y suponiendo una distribución binomial para cada variable. Siendo a y b dos variables aleatorias dicotómicas en una población, E y A y B los subconjuntos donde se verifican a y b , el índice de similaridad viene dado por la expresión (5) (Lerman, 1981):

$$\partial(a, b) = \frac{\text{card}(A, B) - \frac{\text{card}(A)\text{card}(A, B)}{n}}{\sqrt{\frac{\text{card}(A)\text{card}(A, B)}{n}}} \quad (5)$$

Las variables a y b tendrán mayor similaridad cuando el número de elementos comunes sea mayor en relación a la frecuencia esperada en caso de independencia y tiene en cuenta el tamaño muestra. En nuestro caso A representa el conjunto de estudiantes que contesta el ítem a y B el que responde el ítem b , (A, B) el conjunto de los que responden correctamente a los dos ítems. La medida de similaridad induce un orden parcial en el conjunto de ítems. El primer paso en el programa es calcular estos índices, que presentamos en las tablas 6A y 6B.

TABLA 6A – Índices de similaridad entre ítems

| | 1.1 | 1.2 | 2.1 | 2.2 | 2.3 | 3 | 4 | 5.1 | 5.2 | 5.3 | 6.1 | 6.2 | 7.1 | 7.2 | 7.3 |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 1.1 | 1.00 | 0.96 | 0.62 | 0.77 | 0.75 | 0.62 | 0.46 | 0.37 | 0.44 | 0.87 | 0.44 | 0.42 | 0.69 | 0.74 | 0.71 |
| 1.2 | 0.96 | 1.00 | 0.72 | 0.87 | 0.78 | 0.63 | 0.82 | 0.42 | 0.49 | 0.91 | 0.69 | 0.58 | 0.93 | 0.95 | 0.96 |
| 2.1 | 0.62 | 0.72 | 1.00 | 1.00 | 1.00 | 0.80 | 0.84 | 0.77 | 0.64 | 0.98 | 0.91 | 0.85 | 0.90 | 0.93 | 0.99 |
| 2.2 | 0.77 | 0.87 | 1.00 | 1.00 | 1.00 | 0.65 | 0.78 | 0.75 | 0.74 | 1.00 | 0.91 | 0.85 | 0.83 | 0.88 | 0.96 |
| 2.3 | 0.75 | 0.78 | 1.00 | 1.00 | 1.00 | 0.68 | 0.84 | 0.79 | 0.67 | 0.97 | 0.92 | 0.83 | 0.89 | 0.91 | 0.97 |
| 3 | 0.62 | 0.63 | 0.80 | 0.65 | 0.68 | 1.00 | 0.67 | 0.51 | 0.40 | 0.71 | 0.28 | 0.27 | 0.67 | 0.66 | 0.68 |
| 4 | 0.46 | 0.82 | 0.84 | 0.78 | 0.84 | 0.67 | 1.00 | 0.69 | 0.73 | 0.73 | 0.19 | 0.16 | 0.79 | 0.89 | 0.91 |
| 5.1 | 0.37 | 0.42 | 0.77 | 0.75 | 0.79 | 0.51 | 0.69 | 1.00 | 1.00 | 0.99 | 0.32 | 0.31 | 0.49 | 0.50 | 0.50 |
| 5.2 | 0.44 | 0.49 | 0.64 | 0.74 | 0.67 | 0.40 | 0.73 | 1.00 | 1.00 | 0.97 | 0.43 | 0.45 | 0.60 | 0.71 | 0.58 |
| 5.3 | 0.87 | 0.91 | 0.98 | 1.00 | 0.97 | 0.71 | 0.73 | 0.99 | 0.97 | 1.00 | 0.79 | 0.74 | 0.83 | 0.85 | 0.86 |
| 6.1 | 0.44 | 0.69 | 0.91 | 0.91 | 0.92 | 0.28 | 0.19 | 0.32 | 0.43 | 0.79 | 1.00 | 1.00 | 0.61 | 0.45 | 0.58 |
| 6.2 | 0.42 | 0.58 | 0.85 | 0.85 | 0.83 | 0.27 | 0.16 | 0.31 | 0.45 | 0.74 | 1.00 | 1.00 | 0.57 | 0.39 | 0.50 |
| 7.1 | 0.69 | 0.93 | 0.90 | 0.83 | 0.89 | 0.67 | 0.79 | 0.49 | 0.60 | 0.83 | 0.61 | 0.57 | 1.00 | 1.00 | 1.00 |
| 7.2 | 0.74 | 0.95 | 0.93 | 0.88 | 0.91 | 0.66 | 0.89 | 0.50 | 0.71 | 0.85 | 0.45 | 0.39 | 1.00 | 1.00 | 1.00 |
| 7.3 | 0.71 | 0.96 | 0.99 | 0.96 | 0.97 | 0.68 | 0.91 | 0.50 | 0.58 | 0.86 | 0.58 | 0.50 | 1.00 | 1.00 | 1.00 |
| 8 | 0.53 | 0.55 | 0.60 | 0.65 | 0.69 | 0.44 | 0.67 | 0.59 | 0.52 | 0.48 | 0.33 | 0.31 | 0.48 | 0.56 | 0.51 |
| 9.1 | 0.65 | 0.75 | 0.92 | 0.95 | 0.93 | 0.59 | 0.75 | 0.77 | 0.79 | 0.89 | 0.68 | 0.72 | 0.85 | 0.93 | 0.83 |
| 9.2 | 0.80 | 0.69 | 0.92 | 0.90 | 0.97 | 0.39 | 0.53 | 0.94 | 0.99 | 0.93 | 0.83 | 0.82 | 0.93 | 0.93 | 0.86 |
| 10.1 | 0.70 | 0.26 | 0.78 | 0.94 | 0.69 | 0.96 | 0.71 | 0.78 | 0.72 | 0.98 | 0.02 | 0.04 | 0.76 | 0.82 | 0.76 |
| 10.2 | 0.35 | 0.41 | 0.24 | 0.09 | 0.18 | 0.35 | 0.36 | 0.26 | 0.23 | 0.07 | 0.33 | 0.44 | 0.46 | 0.40 | 0.43 |
| 10.3 | 0.43 | 0.46 | 0.55 | 0.68 | 0.32 | 0.67 | 0.61 | 0.34 | 0.51 | 0.69 | 0.31 | 0.36 | 0.56 | 0.68 | 0.75 |
| 11 | 0.46 | 0.21 | 0.16 | 0.15 | 0.09 | 0.19 | 0.36 | 0.86 | 0.76 | 0.60 | 0.27 | 0.21 | 0.65 | 0.56 | 0.68 |
| 12 | 0.60 | 0.41 | 0.86 | 0.91 | 0.84 | 0.51 | 0.53 | 0.56 | 0.79 | 1.00 | 0.20 | 0.16 | 0.86 | 0.86 | 0.98 |
| 13 | 0.53 | 0.36 | 0.12 | 0.48 | 0.09 | 0.53 | 0.18 | 0.45 | 0.36 | 0.80 | 0.43 | 0.47 | 0.67 | 0.64 | 0.77 |
| 14 | 0.69 | 0.44 | 0.27 | 0.30 | 0.08 | 0.39 | 0.52 | 0.59 | 0.44 | 0.74 | 0.24 | 0.23 | 0.79 | 0.85 | 0.79 |
| 15 | 0.87 | 0.55 | 0.71 | 0.88 | 0.66 | 0.80 | 0.58 | 0.81 | 0.83 | 0.92 | 0.24 | 0.22 | 0.80 | 0.86 | 0.91 |
| 16 | 0.86 | 0.48 | 0.59 | 0.76 | 0.89 | 0.79 | 0.52 | 0.54 | 0.36 | 0.53 | 0.55 | 0.41 | 0.68 | 0.73 | 0.74 |

TABLA 6B – *Índices de similitud entre ítems*

| | 8 | 9.1 | 9.2 | 10.1 | 10.2 | 10.3 | 11 | 12 | 13 | 14 | 15 | 16 |
|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 1.1 | 0.53 | 0.65 | 0.80 | 0.70 | 0.35 | 0.43 | 0.46 | 0.60 | 0.53 | 0.69 | 0.87 | 0.86 |
| 1.2 | 0.55 | 0.75 | 0.69 | 0.26 | 0.41 | 0.46 | 0.21 | 0.41 | 0.36 | 0.44 | 0.55 | 0.48 |
| 2.1 | 0.60 | 0.92 | 0.92 | 0.78 | 0.24 | 0.55 | 0.16 | 0.86 | 0.12 | 0.27 | 0.71 | 0.59 |
| 2.2 | 0.65 | 0.95 | 0.90 | 0.94 | 0.09 | 0.68 | 0.15 | 0.91 | 0.48 | 0.30 | 0.88 | 0.76 |
| 2.3 | 0.69 | 0.93 | 0.97 | 0.69 | 0.18 | 0.32 | 0.09 | 0.84 | 0.09 | 0.08 | 0.66 | 0.89 |
| 3 | 0.44 | 0.59 | 0.39 | 0.96 | 0.35 | 0.67 | 0.19 | 0.51 | 0.53 | 0.39 | 0.80 | 0.79 |
| 4 | 0.67 | 0.75 | 0.53 | 0.71 | 0.36 | 0.61 | 0.36 | 0.53 | 0.18 | 0.52 | 0.58 | 0.52 |
| 5.1 | 0.59 | 0.77 | 0.94 | 0.78 | 0.26 | 0.34 | 0.86 | 0.56 | 0.45 | 0.59 | 0.81 | 0.54 |
| 5.2 | 0.52 | 0.79 | 0.99 | 0.72 | 0.23 | 0.51 | 0.76 | 0.79 | 0.36 | 0.44 | 0.83 | 0.36 |
| 5.3 | 0.48 | 0.89 | 0.93 | 0.98 | 0.07 | 0.69 | 0.60 | 1.00 | 0.80 | 0.74 | 0.92 | 0.53 |
| 6.1 | 0.33 | 0.68 | 0.83 | 0.02 | 0.33 | 0.31 | 0.27 | 0.20 | 0.43 | 0.24 | 0.24 | 0.55 |
| 6.2 | 0.31 | 0.72 | 0.82 | 0.04 | 0.44 | 0.36 | 0.21 | 0.16 | 0.47 | 0.23 | 0.22 | 0.41 |
| 7.1 | 0.48 | 0.85 | 0.93 | 0.76 | 0.46 | 0.56 | 0.65 | 0.86 | 0.67 | 0.79 | 0.80 | 0.68 |
| 7.2 | 0.56 | 0.93 | 0.93 | 0.82 | 0.40 | 0.68 | 0.56 | 0.86 | 0.64 | 0.85 | 0.86 | 0.73 |
| 7.3 | 0.51 | 0.83 | 0.86 | 0.76 | 0.43 | 0.75 | 0.68 | 0.98 | 0.77 | 0.79 | 0.91 | 0.74 |
| 8 | 1.00 | 0.69 | 0.65 | 0.33 | 0.55 | 0.68 | 0.60 | 0.60 | 0.41 | 0.57 | 0.57 | 0.75 |
| 9.1 | 0.69 | 1.00 | 1.00 | 0.84 | 0.41 | 0.92 | 0.36 | 0.52 | 0.55 | 0.50 | 0.89 | 0.81 |
| 9.2 | 0.65 | 1.00 | 1.00 | 0.90 | 0.58 | 0.83 | 0.83 | 0.80 | 0.74 | 0.46 | 0.86 | 0.88 |
| 10.1 | 0.33 | 0.84 | 0.90 | 1.00 | 0.84 | 1.00 | 0.87 | 0.56 | 0.97 | 0.93 | 1.00 | 0.97 |
| 10.2 | 0.55 | 0.41 | 0.58 | 0.84 | 1.00 | 1.00 | 0.83 | 0.75 | 0.84 | 0.71 | 0.77 | 0.45 |
| 10.3 | 0.68 | 0.92 | 0.83 | 1.00 | 1.00 | 1.00 | 0.85 | 0.97 | 1.00 | 0.74 | 1.00 | 0.95 |
| 11 | 0.60 | 0.36 | 0.83 | 0.87 | 0.83 | 0.85 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 |
| 12 | 0.60 | 0.52 | 0.80 | 0.56 | 0.75 | 0.97 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.88 |
| 13 | 0.41 | 0.55 | 0.74 | 0.97 | 0.84 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 14 | 0.57 | 0.50 | 0.46 | 0.93 | 0.71 | 0.74 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 15 | 0.57 | 0.89 | 0.86 | 1.00 | 0.77 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 16 | 0.75 | 0.81 | 0.88 | 0.97 | 0.45 | 0.95 | 0.98 | 0.88 | 1.00 | 1.00 | 1.00 | 1.00 |

Una vez obtenidos los índices, se calcula una similitud entre clases: *A* y *B*, el algoritmo de clasificación jerárquica se construye considerando la mayor proximidad entre elementos de una clase y la mayor distancia entre clases separadas (Tabla 7).

TABLA 7A – *Índices de similitud en análisis jerárquico según pasos en la clasificación*

| Paso | Nodos que se unen | Similitud |
|------|-------------------|-----------|
| 1 | (2.1 2.2) | 1 |
| 2 | ((2.1 2.2) 2.3) | 1 |
| 3 | (5.1 5.2) | 1 |
| 4 | (6.1 6.2) | 1 |
| 5 | (12 13) | 1 |
| 6 | (7.2 7.3) | 1 |
| 7 | (15 16) | 1 |

TABLA 7B– Índices de similaridad en análisis jerárquico según pasos en la clasificación

| Paso | Nodos que se unen | Similaridad |
|------|---|-------------|
| 8 | ((12 13) 14) | 1 |
| 9 | (7.1 (7.2 7.3)) | 1 |
| 10 | (9.1 9.2) | 0.999999 |
| 11 | (10.2 10.3) | 0.999997 |
| 12 | ((12 13) 14) (15 16) | 0.999996 |
| 13 | (5.3 (((12 13) 14) (15 16))) | 0.99981 |
| 14 | (10.1 (10.2 10.3)) | 0.999694 |
| 15 | ((5.3 (((12 13) 14) (15 16))) 11) | 0.997586 |
| 16 | ((2.1 2.2) 2.3) ((5.3 (((12 13) 14) (15 16))) 11)) | 0.977147 |
| 17 | (1.1 1.2) | 0.960553 |
| 18 | ((5.1 5.2) (9.1 9.2)) | 0.956171 |
| 19 | ((2.1 2.2) 2.3) ((5.3 (((12 13) 14) (15 16))) 11)) (10.1 (10.2 10.3)) | 0.936159 |
| 20 | ((1.1 1.2) (7.1 (7.2 7.3))) | 0.769849 |
| 21 | (3 4) | 0.673794 |
| 22 | ((2.1 2.2) 2.3) ((5.3 (((12 13) 14) (15 16))) 11)) (10.1 (10.2 10.3)) ((5.1 5.2) (9.1 9.2)) | 0.551825 |
| 23 | (3 4) 8) | 0.447176 |
| 24 | ((1.1 1.2) (7.1 (7.2 7.3))) (((2.1 2.2) 2.3) ((5.3 (((12 13) 14) (15 16))) 11)) (10.1 (10.2 10.3)) ((5.1 5.2) (9.1 9.2))) | 0.342608 |
| 25 | ((1.1 1.2) (7.1 (7.2 7.3))) (((2.1 2.2) 2.3) ((5.3 (((12 13) 14) (15 16))) 11)) (10.1 (10.2 10.3)) ((5.1 5.2) (9.1 9.2))) ((3 4) 8)) | 0.0727032 |
| 26 | ((1.1 1.2) (7.1 (7.2 7.3))) (((2.1 2.2) 2.3) ((5.3 (((12 13) 14) (15 16))) 11)) (10.1 (10.2 10.3)) ((5.1 5.2) (9.1 9.2))) ((3 4) 8))(6.16.2)) | 0.0171955 |

El programa CHIC proporciona también una prueba de significación de las clases obtenidas, que se calcula teniendo en cuenta no sólo la intensidad de la similaridad, sino el número total de sujetos que da una respuesta correcta al conjunto de ítems incluidos en el grupo. Obtuvimos los siguientes grupos:

- Grupo 1: Ítems 1.1 y 1.2. (significado de la media como reparto equitativo y determinar una distribución dada la media), que se unen a su vez con los ítems 7.1, 7.2 y 7.3, también relacionados con hallar una distribución dada la media y el efecto del cero sobre el cálculo de la media.
- Grupo 2: formado por varios subgrupos:
 - Grupo 2.1: Ítems 2.1, 2.2 (cálculo de la media ponderada) y 2.3 (media de la suma de dos variables). Este grupo tiene una alta similaridad, pues el cálculo de la media ponderada se trata de un punto en que fallan muchos estudiantes, según ha mostrado la investigación de Pollatsek, Lima y Well (1981), y en nuestro caso dio un alto índice de errores. La tercera parte del ítem también está relacionada con las anteriores, aunque se tuvo en cuenta su solución correcta independientemente de que el estudiante hubiese calculado correctamente la media en los pasos anteriores.

- Grupo 2.2: Ítems 12 (efecto del valor atípico sobre los promedios) y 13 (mejor representante en caso de distribución no simétrica), que se unen a su vez con el 14 (distribución bimodal) y posteriormente con los ítems 15 (cambio de escala) y 16 (interpretación de promedios en un gráfico conjunto de dos distribuciones). La interpretación de este subconjunto de ítems depende de la comprensión gráfica a nivel intermedio (extracción de tendencias en los datos, según Curcio, 1999). Esta habilidad de interpretación de gráficos ocasiona que los mismos estudiantes respondan correcta o incorrectamente a este conjunto de ítems de gran similitud. Los dos subgrupos anteriores se unen con el 5.3 (elegir el promedio que mejor representa los datos) y el 11 (significado de la media y mediana e interpretación de mediana y moda en un gráfico).
- Grupo 2.3: Ítems 10.1, 10.2 y 10.3 (cálculo de media, mediana y moda a partir de una tabla). Este es un conjunto de ítems muy específico, pues se trata de un procedimiento enseñado habitualmente en la clase, pero que algunos alumnos olvidan o no comprenden, fallando al aplicarlo. Como vemos en la Tabla 5.6.1.3, la similitud entre estos ítems es muy alta, de modo que los alumnos o bien son capaces de aplicar correctamente todas las fórmulas de cálculo, o fallan en todas ellas.
- Grupo 2.4: Ítems 5.1 y 5.2 (cálculo de la mediana con un número par e impar de datos); los alumnos que resuelven uno de estos algoritmos, también lo hacen con el otro, y que se unen con los 9.1 y 9.2 (cálculo de media y mediana a partir de un gráfico), que así mismo, son habilidades relacionadas. En definitiva, este grupo representaría la habilidad de cálculo de la mediana, principalmente.
- Grupo 2.5: Ítems 3 (suma de desviaciones a la media), 4 (media como valor dentro del recorrido) y 8 (media como mejor estimación). En este caso la similitud es mucho menor que en los anteriores, por lo que los estudiantes podrían acertar alguno de estos ítems y fallar en otros. Resaltamos el hecho que se trata en todo caso de comprensión de propiedades de la media.
- Grupo 2.6: Ítems 6.1 y 6.2 (cálculo de mediana en datos ordinales). Muy relacionados entre sí y completamente separados del resto, por lo que la capacidad de trabajo con datos ordinales no se relaciona con el resto de competencias medidas en el cuestionario.

Presentamos también, el dendrograma que muestra los grupos formados de ítems donde los mismos alumnos dan contestaciones similares (bien correctas, bien incorrectas). Es decir, se trata de conocimientos relacionados entre sí y separados de los otros grupos de ítems (Figura 1). Observamos una estructura muy compleja con numerosos grupos, lo que indica componentes diferenciados en el significado de las medidas de tendencia central.

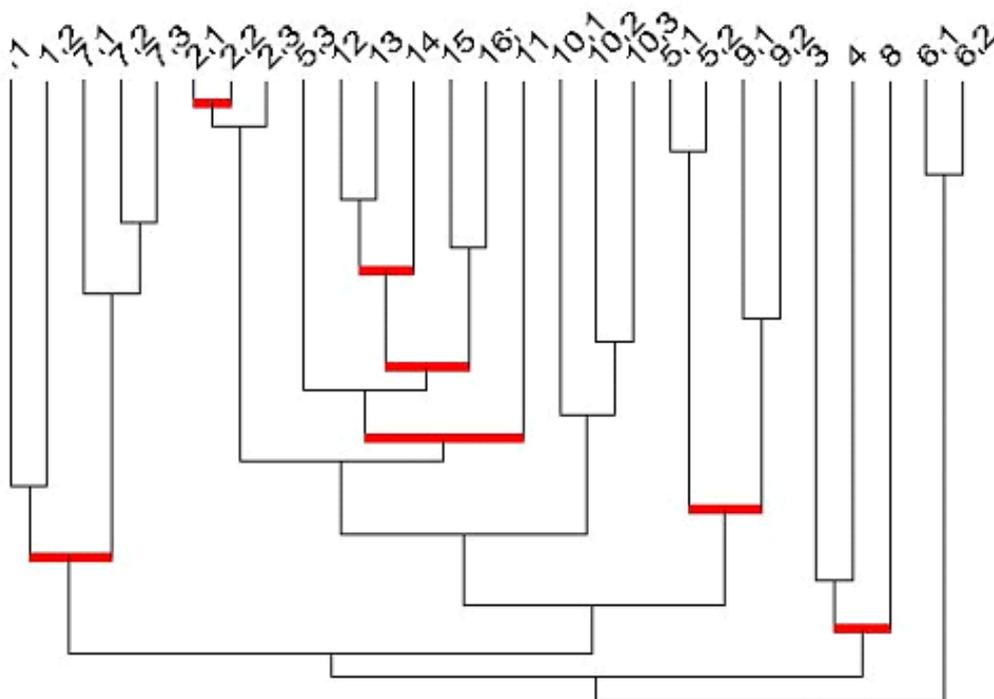


FIGURA 1 – Árbol de similaridad con todas las variables, método clásico, ley binomial

3.2 Grafo Implicativo

El análisis clúster presentado considera una medida de asociación simétrica en las variables, es decir, se supone que si un estudiante responde a un ítem también responderá al asociado con él, pero no tiene en cuenta la dificultad relativa de cada ítem. Una situación más plausible es pensar que, aunque dos ítems estén relacionados, si uno es más difícil, la respuesta a este ítem facilita que también se acierte en el segundo.

Para estudiar este punto hemos llevado a cabo un análisis implicativo entre ítems, que proporciona un estudio de la implicación (no simétrica) entre el conjunto de ítems, es decir, se trata de ver si la respuesta correcta al ítem a implica la respuesta correcta al b (donde la respuesta correcta a b puede o puede que no implique la respuesta a a).

Siendo a y b dos variables aleatorias dicotómicas en una población E y A y B los subconjuntos donde se verifican a y b . El índice de implicación de Gras (Gras, 1996) (Gras, Régnier, Guillet 2009) viene dado por la expresión (6):

$$q(a, \bar{b}) = \left[\frac{\text{card}(A \cap \bar{B}) - \frac{\text{card}(A)\text{card}(\bar{B})}{n}}{\sqrt{\frac{\text{card}(A)\text{card}(\bar{B})}{n}}} \right] \quad (6)$$

Como antes, a y b representan dos ítems, A y B el conjunto de alumnos que lo responden correctamente, \bar{B} el conjunto de alumnos que comete algún error en el ítem b . El programa también calcula la significación estadística del índice, que sigue una distribución $N(0,1)$. En nuestro caso, las variables aleatorias son las respuestas

(correcta- incorrecta) a cada ítem, con lo que podemos calcular un total de $\binom{27}{2}$ índices de implicación entre los 27 ítems del cuestionario.

La relación de implicación entre variables establece un preorden asimétrico dentro del conjunto de ítems, que podemos representar en un grafo ordenado. El programa CHIC calcula los índices de implicación entre todos los pares de variables de un conjunto de datos y proporciona un grafo mostrando todas las implicaciones que son significativas hasta el nivel pedido por el usuario, teniendo en cuenta tanto la intensidad de implicación, como el número de sujetos en que se cumple la relación de implicación.

Para analizar la interrelación entre ítems del cuestionario, mostramos los índices de implicación entre ítems de la prueba (Tablas 8A y 8B), así como el grafo implicativo (Figura 2), donde se presentan en rojo las relaciones significativas al 0.99.

TABLA 8A – *Índices de implicación (teoría clásica)*

| | 1.1 | 1.2 | 2.1 | 2.2 | 2.3 | 3 | 4 | 5.1 | 5.2 | 5.3 | 6.1 | 6.2 | 7.1 | 7.2 | 7.3 |
|------|-----|-----|-----|-----|-----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1.1 | 0 | 99 | 58 | 71 | 65 | 66 | 43 | 32 | 43 | 77 | 45 | 44 | 82 | 87 | 80 |
| 1.2 | 100 | 0 | 66 | 81 | 67 | 67 | 95 | 40 | 49 | 80 | 66 | 56 | 100 | 100 | 99 |
| 2.1 | 68 | 80 | 0 | 100 | 100 | 87 | 96 | 84 | 66 | 91 | 87 | 79 | 99 | 100 | 100 |
| 2.2 | 87 | 95 | 100 | 0 | 100 | 70 | 90 | 82 | 76 | 98 | 87 | 79 | 96 | 98 | 100 |
| 2.3 | 84 | 85 | 100 | 100 | 0 | 74 | 96 | 86 | 69 | 89 | 88 | 77 | 99 | 99 | 100 |
| 3 | 68 | 68 | 71 | 62 | 61 | 0 | 79 | 52 | 38 | 64 | 32 | 32 | 80 | 77 | 75 |
| 4 | 44 | 90 | 75 | 72 | 72 | 73 | 0 | 75 | 75 | 65 | 23 | 22 | 93 | 98 | 98 |
| 5.1 | 30 | 39 | 69 | 69 | 68 | 52 | 80 | 0 | 100 | 93 | 35 | 35 | 48 | 51 | 50 |
| 5.2 | 40 | 49 | 60 | 68 | 60 | 36 | 86 | 100 | 0 | 89 | 44 | 46 | 68 | 83 | 62 |
| 5.3 | 96 | 97 | 92 | 99 | 87 | 77 | 85 | 100 | 98 | 0 | 74 | 69 | 96 | 96 | 95 |
| 6.1 | 40 | 76 | 82 | 85 | 79 | 22 | 6 | 26 | 42 | 70 | 0 | 100 | 70 | 42 | 62 |
| 6.2 | 38 | 61 | 75 | 78 | 71 | 21 | 4 | 25 | 45 | 66 | 100 | 0 | 62 | 32 | 50 |
| 7.1 | 78 | 98 | 80 | 77 | 76 | 72 | 92 | 49 | 61 | 73 | 59 | 55 | 0 | 100 | 100 |
| 7.2 | 85 | 99 | 84 | 82 | 78 | 71 | 99 | 51 | 73 | 75 | 46 | 42 | 100 | 0 | 100 |
| 7.3 | 81 | 99 | 93 | 91 | 87 | 73 | 99 | 50 | 59 | 76 | 56 | 50 | 100 | 100 | 0 |
| 8 | 54 | 57 | 56 | 61 | 61 | 42 | 78 | 62 | 52 | 49 | 36 | 35 | 46 | 60 | 52 |
| 9.1 | 72 | 82 | 83 | 89 | 80 | 62 | 88 | 84 | 82 | 78 | 65 | 68 | 97 | 99 | 92 |
| 9.2 | 89 | 75 | 82 | 83 | 86 | 35 | 51 | 98 | 99 | 83 | 78 | 76 | 100 | 100 | 94 |
| 10.1 | 77 | 19 | 70 | 89 | 61 | 99 | 82 | 84 | 75 | 91 | 5 | 8 | 90 | 93 | 84 |
| 10.2 | 28 | 38 | 32 | 15 | 30 | 31 | 27 | 19 | 21 | 17 | 36 | 45 | 38 | 33 | 40 |
| 10.3 | 39 | 44 | 53 | 64 | 39 | 72 | 69 | 29 | 51 | 62 | 35 | 39 | 62 | 78 | 85 |
| 11 | 43 | 13 | 25 | 21 | 22 | 12 | 27 | 93 | 78 | 56 | 31 | 27 | 73 | 57 | 75 |
| 12 | 61 | 37 | 76 | 84 | 71 | 51 | 52 | 59 | 82 | 99 | 25 | 21 | 98 | 97 | 100 |
| 13 | 55 | 31 | 22 | 48 | 22 | 54 | 5 | 43 | 35 | 71 | 44 | 47 | 77 | 73 | 86 |
| 14 | 77 | 41 | 34 | 35 | 21 | 36 | 53 | 63 | 43 | 66 | 28 | 28 | 93 | 96 | 89 |
| 15 | 96 | 57 | 64 | 82 | 59 | 87 | 63 | 89 | 86 | 81 | 28 | 28 | 95 | 97 | 98 |
| 16 | 95 | 48 | 56 | 71 | 76 | 86 | 53 | 55 | 34 | 52 | 54 | 43 | 79 | 84 | 83 |

TABLA 8B - *Índices de implicación (teoría clásica)*

| | 8 | 9.1 | 9.2 | 10.1 | 10.2 | 10.3 | 11 | 12 | 13 | 14 | 15 | 16 |
|------|----|-----|-----|------|------|------|----|-----|-----|-----|-----|-----|
| 1.1 | 56 | 68 | 70 | 62 | 40 | 42 | 47 | 55 | 52 | 63 | 86 | 77 |
| 1.2 | 62 | 79 | 62 | 36 | 44 | 45 | 30 | 45 | 40 | 46 | 55 | 49 |
| 2.1 | 68 | 96 | 81 | 66 | 32 | 55 | 26 | 72 | 21 | 34 | 70 | 56 |
| 2.2 | 80 | 98 | 79 | 81 | 19 | 71 | 25 | 76 | 49 | 36 | 87 | 69 |
| 2.3 | 87 | 97 | 88 | 60 | 28 | 30 | 19 | 70 | 17 | 17 | 65 | 80 |
| 3 | 36 | 62 | 43 | 83 | 40 | 69 | 28 | 50 | 52 | 43 | 79 | 71 |
| 4 | 86 | 79 | 52 | 62 | 41 | 62 | 41 | 52 | 26 | 51 | 57 | 51 |
| 5.1 | 71 | 81 | 84 | 66 | 34 | 32 | 76 | 53 | 46 | 56 | 80 | 53 |
| 5.2 | 55 | 84 | 93 | 63 | 32 | 51 | 68 | 67 | 40 | 46 | 82 | 40 |
| 5.3 | 41 | 93 | 83 | 88 | 17 | 71 | 56 | 98 | 72 | 66 | 91 | 52 |
| 6.1 | 13 | 72 | 73 | 14 | 39 | 29 | 34 | 33 | 45 | 32 | 25 | 53 |
| 6.2 | 10 | 77 | 72 | 16 | 46 | 35 | 30 | 29 | 48 | 31 | 23 | 44 |
| 7.1 | 45 | 90 | 83 | 65 | 47 | 57 | 60 | 72 | 62 | 71 | 79 | 63 |
| 7.2 | 64 | 96 | 83 | 69 | 43 | 70 | 54 | 73 | 60 | 76 | 85 | 66 |
| 7.3 | 53 | 88 | 75 | 65 | 46 | 78 | 62 | 86 | 70 | 71 | 90 | 67 |
| 8 | 0 | 73 | 60 | 40 | 54 | 70 | 57 | 56 | 44 | 55 | 56 | 68 |
| 9.1 | 88 | 0 | 100 | 71 | 44 | 94 | 41 | 51 | 53 | 50 | 88 | 72 |
| 9.2 | 79 | 100 | 0 | 76 | 55 | 86 | 74 | 68 | 67 | 47 | 85 | 79 |
| 10.1 | 12 | 89 | 80 | 0 | 74 | 100 | 77 | 53 | 91 | 84 | 99 | 90 |
| 10.2 | 58 | 39 | 55 | 70 | 0 | 100 | 74 | 65 | 76 | 64 | 77 | 47 |
| 10.3 | 86 | 96 | 73 | 98 | 100 | 0 | 75 | 84 | 97 | 66 | 100 | 87 |
| 11 | 70 | 33 | 73 | 73 | 73 | 88 | 0 | 93 | 99 | 98 | 100 | 93 |
| 12 | 69 | 52 | 70 | 53 | 67 | 98 | 96 | 0 | 100 | 100 | 100 | 79 |
| 13 | 25 | 56 | 66 | 85 | 74 | 100 | 99 | 100 | 0 | 100 | 100 | 99 |
| 14 | 62 | 50 | 47 | 79 | 64 | 76 | 98 | 100 | 100 | 0 | 100 | 99 |
| 15 | 62 | 94 | 75 | 92 | 69 | 100 | 99 | 98 | 100 | 100 | 0 | 100 |
| 16 | 95 | 86 | 77 | 85 | 47 | 97 | 92 | 74 | 99 | 98 | 100 | 0 |

La implicación entre un ítem y otro se interpreta en el sentido de que si un estudiante es capaz de resolver correctamente un ítem, entonces mejora su probabilidad de resolver correctamente otro implicado por aquél. En este sentido, el árbol implicativo nos proporciona una pauta de posible orden de introducción de los conceptos y procedimientos evaluados por los diferentes ítems.

Hacemos notar que la relación de implicación es asimétrica, indicándose el sentido de la implicación por la dirección de la flecha en el grafo. Por tanto, si estudiamos las relaciones significativas al 0.99, que aparecen en rojo en la Figura 2, la interpretación es que los alumnos resuelven correctamente el ítem.

más complejo que el de media. Por otro lado, el estudiante que calcula la mediana a partir del gráfico también realiza mejor el problema de determinar una distribución (ítem 7.1) y dar un segundo ejemplo (ítem 7.2), posiblemente porque la interpretación correcta de un gráfico implica la comprensión de la idea de distribución representada en el gráfico.

- La comprensión del efecto del cero sobre la media (ítem 7.3) facilita el dar una distribución con media dada (ítem 7.1) y la comprensión de que el valor de la media se sitúa en el centro del recorrido (ítem 4); todos estos ítems se refieren a propiedades formales de la media, con lo que la comprensión de algunas facilita las otras.
- La comprensión del efecto de un valor atípico sobre el cálculo de medidas de tendencia central (ítem 12), facilita elegir el mejor representante entre media y mediana para un ejemplo particular (ítem 5.3) y éste facilita el cálculo correcto de la mediana en conjunto de datos impar (ítem 5.1) y par (ítem 5.2) respectivamente.
- La interpretación correcta de media y mediana en un gráfico (ítem 11) facilita el comprender los casos de distribución binomial (ítem 14) y éste el de dos distribuciones representadas en el mismo gráfico (ítem 16). Éste, a su vez, la interpretación de valor típico en un gráfico (ítem 13) y el cambio de escala (ítem 15). Todos estos ítems están relacionados con el razonamiento estadístico a partir de gráficos y su interpretación a nivel de extracción de tendencias (Curcio, 1989), y a su vez, implican el cálculo correcto de moda (ítem 10.3) y mediana (ítem 10.2) a partir de un gráfico.

En definitiva, el grafo implicativo apoya nuestra hipótesis de que la comprensión de medidas de tendencia central por los estudiantes mexicanos no puede concebirse como un constructo unitario, lo que explica que el análisis factorial haya resultado con tantos factores, así como que la fiabilidad (coeficiente Alfa) tuviese un valor moderado. Por el contrario, el grafo implicativo muestra una jerarquía de conocimientos entrelazados que los alumnos han de conseguir progresivamente y que el profesor debe considerar al planificar la enseñanza del tema.

3.3 Análisis Implicativo Jerárquico

El grafo implicativo, aunque muestra la estructura de interrelaciones, es algo complejo, por lo que sería interesante tratar de dividir el conjunto de ítems en unos pocos grupos interrelacionados entre sí mediante el índice de implicación.

Una vez estudiadas las implicaciones aisladas de unos ítems sobre otros, hemos llevado a cabo un estudio de clasificación implicativa. Se trata de un algoritmo que utiliza las intensidades de implicaciones entre conjuntos de variables como índice no simétrico para estudiar la cohesión interna de algunos subconjuntos de variables. La cohesión de una clase tiene en cuenta la cantidad de información proporcionada por un conjunto de variables, el índice se puede interpretar como cantidad de información que una variable proporciona sobre otra.

Dadas dos variables a y b , en nuestro caso respuestas a dos ítems, y siendo $p = \max(\varphi(a, \bar{b}), \varphi(b, \bar{a}))$ y $H = -p \log_2 p - (1-p) \log_2 (1-p)$, la entropía del conjunto

de variables, el grado de cohesión entre dos variables se define como $c(a, b) = \sqrt{1 - H^2}$ que también varía entre 0 (mínimo) y 1 (máximo). El grado de cohesión de una clase se calcula en cada paso con la fórmula (7) (Gras, Kuntz y Briand, 2001):

$$C(\underline{A}) = \left[\prod_{\substack{i \in \{1, \dots, r-1\} \\ j \in \{2, \dots, r\}, j > i}} c(a_i, a_j) \right]^{\frac{2}{r(r-1)}} \quad (7)$$

Por último, dados \underline{A} y \underline{B} dos conjuntos de variables, formados por variables a_i y b_j y $C(\underline{A})$ y $C(\underline{B})$, sus índices de cohesión, la intensidad de implicación de la clase \underline{A} sobre la clase \underline{B} se define con (8) (Couturier, 2001):

$$\psi(\underline{A}, \underline{B}) = \left[\sup_{i \in \{1, \dots, r\}, j \in \{1, \dots, s\}} \varphi(a_i, \bar{b}_j) \right]^{rs} [C(\underline{A}) \cdot C(\underline{B})]^{1/2} \quad (8)$$

El programa CHIC calcula el nivel de significación de los diferentes nodos en una jerarquía implicativa, así como las contribuciones de los sujetos. El algoritmo forma las clases teniendo en cuenta los siguientes criterios: 1) la cohesión máxima dentro de cada clase; 2) el mayor grado de implicación entre una clase y otra que es implicada por ella.

Completamos el estudio con la determinación de una jerarquía implicativa en el conjunto de variables, presentando en la Tabla 9 los coeficientes de cohesión en los distintos pasos del procedimiento y en la Figura 3 el árbol de cohesión implicativa.

TABLA 9 - *Índices de similaridad en análisis jerárquico según pasos en la clasificación*

| Nivel | Nodos que se unen | Cohesión |
|-------|--|----------|
| 1 | (2.1 2.2) | 1 |
| 2 | ((2.1 2.2) 2.3) | 1 |
| 3 | (5.1 5.2) | 1 |
| 4 | (6.1 6.2) | 1 |
| 5 | (7.1 7.2) | 1 |
| 6 | ((7.1 7.2) 7.3) | 1 |
| 7 | (9.2 9.1) | 1 |
| 8 | (10.2 10.3) | 1 |
| 9 | (16 15) | 1 |
| 10 | (12 13) | 1 |
| 11 | ((12 13) 14) | 1 |
| 12 | (1.2 ((7.1 7.2) 7.3)) | 1 |
| 13 | (5.3 (5.1 5.2)) | 0.998 |
| 14 | (10.1 3) | 0.998 |
| 15 | ((((12 13) 14) 11) | 0.992 |
| 16 | ((1.2 ((7.1 7.2) 7.3)) 4) | 0.986 |
| 17 | ((16 15) .1) | 0.977 |
| 18 | ((((2.1 2.2) 2.3) ((1.2 ((7.1 7.2) 7.3)) 4)) | 0.964 |
| 19 | ((5.3 (5.1 5.2)) (9.2 9.1)) | 0.88 |
| 20 | (((((12 13) 14) 11) (10.2 10.3)) | 0.771 |
| 21 | ((10.1 3) ((16 15) 1.1)) | 0.771 |

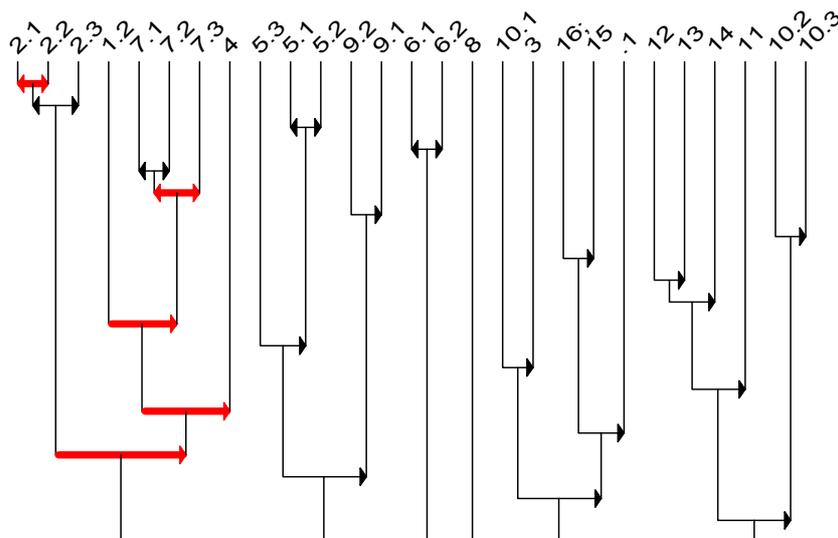


FIGURA 3 - Árbol de cohesión implicativa

Observamos cinco grandes grupos de ítems:

- Primer grupo: *Cálculo avanzado de la media y comprensión procedimental*. Ítems 2.1, 2.2 (*cálculo de la media ponderada*) y 2.3 (*media de la suma de dos variables*), que se implican todos ellos entre sí; 1.2 (*determinar una distribución dada la media*), que implica a los ítems 7.1, 7.2, 7.3 (*hallar una distribución dada la media y el efecto del cero sobre el cálculo de la media*), que se implican entre sí e implican al ítem 4 (*media como valor dentro del recorrido*).
- Segundo grupo: *Cálculo de la mediana*. Ítem 5.3, que implica a los ítems 5.1 y 5.2 (*cálculo de la mediana con un número par e impar de datos*), que se implican entre sí e implican al 9.2 (*cálculo de mediana a partir de un gráfico*), que a su vez, implica al 9.1 (*cálculo de media partir de un gráfico*).
- Tercer grupo: *Mediana en datos ordinales*. Ítems 6.1 y 6.2 (*cálculo de mediana en datos ordinales*), que se implican entre sí y completamente separados del resto.
- Cuarto grupo: *Comprensión conceptual de la media*. Ítem 10.1 (*cálculo de la media en una tabla*), que implica al ítem 3 (*suma de desviaciones a la media*), que implican al grupo formado por los ítems 16 (*interpretación de promedios en gráfico conjunto de dos distribuciones*) y el 15 (*cambio de escala*), y que implican al 1.1 (*significado de la media como reparto equitativo*).
- Quinto grupo: *Representante de un conjunto de datos*. Ítem 12 (*efecto del valor atípico sobre los promedios*), que implica al ítem 13 (*mejor representante en caso de distribución no simétrica*), que se unen a su vez con el 14 (*distribución bimodal*) y posteriormente con el 11 (*significado de la media y mediana e interpretación de mediana y moda en un gráfico*), y 10.2, que implica al 10.3 (*cálculo de media, mediana y moda a partir de una tabla*).

- El ítem 8 (*media como mejor estimación*) aparece separado del resto.

3.4 Conclusiones

Los resultados del estudio de análisis cluster e implicativo muestran la estructura compleja del cuestionario e indican que la comprensión de ciertos elementos de significado de las medidas de tendencia central no se relaciona con los de otros. Observamos cómo los grupos que se relacionan con la comprensión conceptual (de definiciones o de propiedades) aparecen separados de los relacionados con aspectos procedimentales y también aparecen como separados los grupos de ítems relacionados con la mediana (grupos dos y tres) de los relacionados con la media.

De este modo se sugiere que la comprensión de la mediana y de la media no está relacionada en estos estudiantes.

Otra aportación de este estudio es la validación del cuestionario, pues hemos argumentado, al igual que hizo Cobo (2003) su validez de contenido, pero se ha añadido el estudio de la validez discriminante y de constructo que Cobo (2003) no llevó a cabo. Hemos repetido también los cálculos de índices de fiabilidad y generalizabilidad, pero ahora se completan con el coeficiente Theta basado en el análisis factorial, más adecuado para cuestionarios multidimensionales como este. También se proporcionan estimaciones bayesianas de los índices de dificultad del cuestionario, muy apropiadas para situaciones con información previa, como la nuestra.

Respecto a la validez de constructo, se ha repetido el análisis clúster que realizó Cobo (2003), añadiéndose ahora el análisis implicativo para confirmar la estructura de las respuestas al cuestionario.

Referencias

- [1] Barbero, M. (2003), *Psicometría II. Métodos de elaboración de escalas*. Madrid, UNED.
- [2] Batanero, C. (2000), Significado y comprensión de las medidas de tendencia central. *UNO*, **25**, 41-58.
- [3] Batanero, C. y Díaz, M. C. (2005), Análisis del proceso de construcción de un cuestionario sobre probabilidad condicional. Reflexiones desde el marco de la TFS, En A. Contreras (Ed.), *Investigación en Didáctica de las Matemáticas*, 13-36, Universidad de Jaén.
- [4] Bisquerra, R. (1989), *Métodos de investigación educativa*, Barcelona, CEAC.
- [5] Carmines, E. G. y Zeller, R. A. (1979), *Reliability and validity assessment*, Londres, Sage University Paper.
- [6] Carmona, J. (2004), Una revisión de las evidencias de fiabilidad y validez de los cuestionarios de actitudes y ansiedad hacia la estadística, *Statistics Education Research Journal*. **3**(1), <http://www.stat.auckland.ac.nz/~iase/serj/>.
- [7] Cobo, B. (2003), *Significado de las medidas de posición central para los estudiantes de secundaria*, Tesis Doctoral, Universidad de Granada.

- [8] Cobo, B. y Batanero, C. (2000), La mediana ¿Un concepto sencillo en la enseñanza secundaria?, *UNO*, **23**, 85-96.
- [9] Couturier, R. y Gras, R. (2005), CHIC: Traitement de données avec l'analyse implicative, En G. Ritschard y C. Djeraba (Eds.), *Journées d'extraction et gestion des connaissances (EGC'2005)*, Vol. **2**, 679-684, Universidad de Lille.
- [10] Curcio, F. R. (1989), *Developing graph comprehension*, Reston, VA: N.C.T.M.
- [11] Dane, F. C. (1990), *Research methods*, Thompson, Pacific Grow, CA.
- [12] Díaz, C. (2007), *Viabilidad de la inferencia bayesiana en el análisis de datos en psicología*, Tesis Doctoral, Universidad de Granada.
- [13] Díaz, C., Batanero, C. y Cobo, B. (2003), Fiabilidad y generalizabilidad, Aplicaciones en evaluación educativa, *Números*, **54**, 3-21.
- [14] Dunn, O. J. y Clarck, V. A. (1987), *Applied statistics: Analysis of variance and regression*, Nueva York, John Wiley.
- [15] Feldt, L. S. y Brennan, R. L. (1991), Reliability, En R. L. Linn (Ed.), *Educational measurement*, 105-146, Nueva York, MacMillan.
- [16] Garfield, J. B. y Konold, C. (1992), *Statistical reasoning assessment. Part 2: Statistics in context*, Minnesota, MN, National Science Foundation.
- [17] Gattuso, L. y Mary, C. (1996), Development of concepts of the arithmetic average from high school to University, *Proceedings of the 20th Conference of the International Group for the Psychology of Mathematics Education*, Vol. **1**, 401-408, Universidad de Valencia.
- [18] Godino, J. D. (1996), Relaciones entre la investigación en didáctica de las matemáticas y la práctica de la enseñanza, En L. Puig y J. Calderón (Eds.) *Investigación y Didáctica de las Matemáticas*, 119-128, Madrid: CIDE.
- [19] Gras, R. (1996), *L'implication statistique: nouvelle méthode exploratoire de données applications a la didactique*, Grenoble: La Pensée Sauvage.
- [20] Gras, R., Kuntz P. y Briand, H. (2001). Les fondements de l'analyse statistique implicative et quelques prolongements pour la fouille de données, *Mathématiques et Sciences Humaines*, 154-155, 9-29.
- [21] Gras R., Régnier J.-C., Guillet F. (Eds) (2009) *Analyse Statistique Implicative. Une méthode d'analyse de données pour la recherche de causalités*. RNTI-E-16 Toulouse: Cépadues
- [22] León, O. G. y Montero, I. (2002), *Métodos de investigación en psicología y educación*, Madrid, McGraw-Hill.
- [23] Lerman, I. C. (1981), *Classification et analyse ordinaire des données*, París, Dunod.
- [24] Martínez Arias, R. (1995), *Psicometría: teoría de los tests psicológicos y educativos*, Madrid, Síntesis.
- [25] Mayén, S. (2009), *Comprensión de medidas de tendencia central en estudiantes mexicanos de educación secundaria y bachillerato*, Tesis Doctoral, Universidad de Granada.

- [26] Mayén, S. (2006), *Comprensión de medidas de posición central en estudiantes mexicanos de Bachillerato*, Memoria de Tercer Ciclo, Universidad de Granada.
- [27] Morales, P. (1988), *Medición de actitudes en psicología y educación*, San Sebastián, Universidad de Comillas.
- [28] Messick, S. (1989), Validity. En R. L. Linn (Ed.), *Educational Measurement*. 3ª. Edición, 13-103, Nueva York: Collier Macmillan.
- [29] Muñoz, J. (1994), *Teoría clásica de los tests*, Madrid, Pirámide.
- [30] Pollatsek, A., Lima, S. y Well, A. D. (1981), Concept or computation: Students' understanding of the mean, *Educational Studies in Mathematics*, **12**, 191-204.
- [31] Santisteban, C. (1990), *Psicometría. Teoría y práctica en la construcción de tests*, Madrid, Norma.
- [32] Sax, G. (1989), *Principles of educational and psychological measurement and evaluation*, Belmont, CA, Wadsworth.
- [33] Thorndike, R. L. (1989), *Psicometría aplicada*, México, Limusa.
- [34] Tormo, C. (1993), *Estudio sobre cuatro propiedades de la media aritmética en alumnos de 12 a 15 años*, Memoria de Tercer Ciclo, Universidad de Valencia.
- [35] Watson, J. (1997), Assessing statistical literacy through the use of media surveys, En I. Gal y J. B. Garfield (Eds.), *The assessment challenge in statistics education*, 107-121, Amsterdam, IOS Press.
- [36] Zawojewski, J. (1986), *The teaching and learning processes of junior high school students under alternative modes of instruction in the measures of central tendency*, Tesis Doctoral, University Northwestern, Evanston, Illinois.

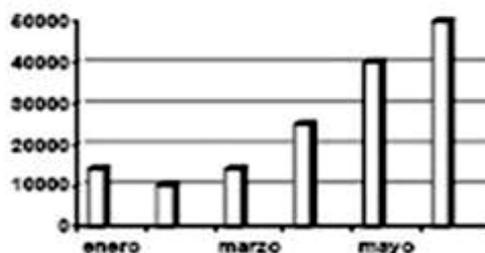
Anexo. Cuestionario

1. Un periódico dice que el número medio de hijos por familia en México es 2.2 hijos por familia. Explica qué significa para ti esta frase.
Se han elegido 10 familias mexicanas y el número medio de hijos entre las 10 familias es de 2.2 hijos por familia. Los García tienen 4 hijos y los Pérez tienen 1 hijo, ¿cuántos hijos podrán tener las otras 8 familias para que la media de hijos en las 10 familias sea 2.2? Explica tu respuesta.
2. María y Pedro dedican una media de 8 horas cada fin de semana a hacer deporte. Otros 8 estudiantes dedican cada fin de semana una media de 4 horas a hacer deporte.
 - a) ¿Cuál es el número medio de horas que hacen deporte cada fin de semana los 10 estudiantes?
María y Pedro dedican además 1 hora cada fin de semana a escuchar música y los otros 8 estudiantes, 3 horas.
 - b) ¿Cuál es el número medio de horas que escuchan música los diez estudiantes?
 - c) ¿Cuál sería el número medio de horas que estos 10 estudiantes dedican cada fin de semana, entre las dos actividades: hacer deporte y escuchar música?

3. Cuatro amigos se reúnen para preparar una cena. Cada uno de ellos trajo harina para hacer la masa de las pizzas. Como querían hacer cuatro pizzas del mismo tamaño, los que habían traído más harina regalaron a los que llevaban menos. ¿La cantidad de harina regalada por los que habían traído mucha fue **mayor, menor o igual** a la recibida por los que habían traído poca? ¿Por qué piensas eso?
4. Tenemos **seis números** y el más grande es el 5. Sumamos estos números y dividimos la suma entre **seis**. El resultado es 4. ¿Te parece posible? ¿Por qué?
5. El peso en kilos de 9 niños es 15, 25, 17, 19, 16, 26, 18, 19, 24. ¿Cuál es el peso del niño **mediano**? ¿Cuál es la **mediana** si incluimos el peso de otro niño que pesa 43 Kg? En este caso, ¿Sería la **media aritmética** un buen representante de los 10 datos? ¿Por qué?
6. Un profesor califica a sus alumnos del siguiente modo: I=Insuficiente, A=Aprobado, N=Notable, S=Sobresaliente. En la siguiente tabla tenemos las notas que ha puesto a dos grupos de alumnos:

| | |
|----------------|--|
| Grupo 1 | I A A N N S S I I I A A A N S S I A A S S S S |
| Grupo 2 | S S I I A N A N I I S N A S I N N |

- a) ¿Qué grupo ha obtenido mejores notas?
- b) ¿Cuál sería el promedio (medida de centralización) más apropiado para representar estos datos?
Explica tu respuesta.
7. Lucía, Juan y Pablo van a una fiesta. Cada uno lleva cierto número de caramelos. Entre todos llevan una media de 11 caramelos por persona.
 - a) ¿Cuántos caramelos ha llevado cada uno? Lucía _____ Juan _____ Pablo _____
 - b) ¿Es la única posibilidad?
Si _____ No _____ Explica cómo has obtenido tus resultados.
Un cuarto chico llega a la fiesta y no lleva ningún caramelo. ¿Cuál es ahora la media de caramelos por chico? Explica tu resultado.
8. Nueve estudiantes han pesado un objeto en la clase de ciencias, usando la misma escala. Los pesos registrados por cada estudiante (en gramos) se muestran a continuación: 6.2, 6.3, 6.0, 15.2, 6.2, 6.1, 6.5, 6.2, 6.1, 6.2. Los estudiantes quieren determinar con la mayor precisión posible el peso real del objeto. ¿Qué harías para calcularlo?
9. Observa el siguiente diagrama de barras que muestra las ventas de bocadillos de la empresa *Bocatta* durante los últimos 6 meses del año pasado:



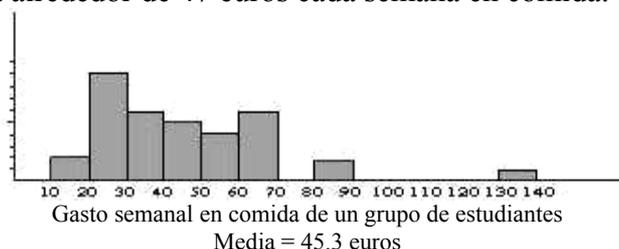
- a) Da un valor aproximado del número medio de bocadillos que se vende al mes.
- b) Da un valor aproximado de la mediana del número de bocadillos que se vendieron por mes.

10. El siguiente conjunto de datos refleja las edades en que contrajeron matrimonio una muestra de 100 mujeres.

| Edad | Frecuencia |
|-------|------------|
| 15-19 | 4 |
| 20-24 | 38 |
| 25-29 | 28 |
| 30-34 | 20 |
| 35-39 | 8 |
| 40-44 | 1 |
| 45-49 | 1 |

¿Cuál es la media, mediana y moda de la edad de estas mujeres? Realiza los cálculos necesarios.

11. Pedro piensa que en el siguiente gráfico, la mediana te dice que una mayoría de estudiantes gasta alrededor de 47 euros cada semana en comida.



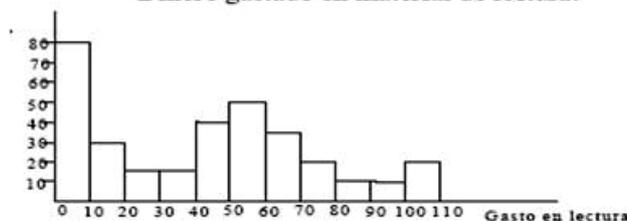
Si _____ No _____ ¿Por qué?

12. Juana piensa que, en el gráfico anterior, el alto valor de 133 euros debería quitarse del conjunto de datos antes de calcular media, mediana y moda.

Si _____ No _____ ¿Por qué?

13. Un grupo de estudiantes hizo este otro histograma del dinero gastado cada semana en material de lectura.

Dinero gastado en material de lectura:



Media = 29 Euros. Mediana = 25 Euros

Francisco dijo que es difícil describir el dinero típico gastado en material de lectura en el gráfico anterior, porque la mayor parte de los estudiantes no gastaron nada o muy poco y un segundo grupo gastó entre 50 y 60 euros cada semana. Él piensa que en este caso la media es un indicador pobre del promedio y elige usar la mediana en su lugar para representar el gasto semanal “promedio” en material de lectura.

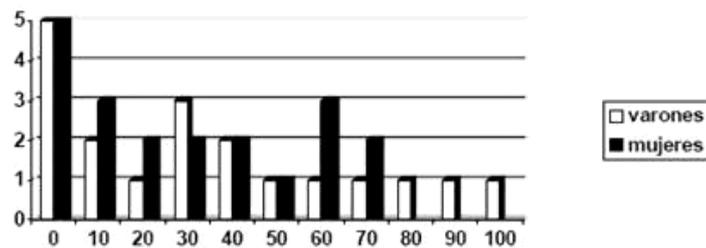
Si _____ No _____ ¿Por qué?

14. Manuel dice que esta distribución tiene dos modas y sugiere que puedes obtener más información de estos datos si los divides en dos subgrupos y calculas el promedio en cada grupo por separado.

Si _____ No _____ ¿Por qué?

15. Lola sugiere hacer el estudio en pesos mexicanos (1€ = \$14.5). ¿Cuál sería en este caso el valor de la media?

16. Antonio quiere investigar las diferencias en los hábitos de gasto en varones y mujeres. Compara las cantidades gastadas en material de lectura en varones y mujeres, construyendo los gráficos siguientes:



Antonio cree que sus gráficos muestran que los varones y mujeres tienden a gastar diferentes cantidades de dinero en material de lectura.

Si _____ No _____ ¿Por qué?