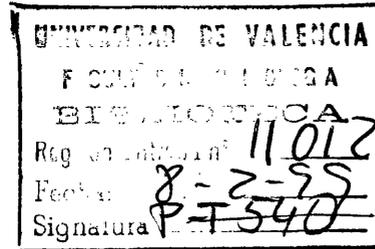


BID.T/503

UNIVERSITAT DE VALÈNCIA

Facultat de Psicologia



ESCALAS DE RESPUESTA Y SISTEMAS DE PUNTUACIÓN: UN ANÁLISIS MEDIANTE MODELOS POLITÓMICOS DE LA TEORÍA DE RESPUESTA AL ÍTEM



TESIS DOCTORAL

Presentada por:

BEGOÑA ESPEJO TORT

Dirigida por:

DR. VICENTE GONZÁLEZ ROMÁ
Profesor Titular de Metodología

Valencia, Julio de 1998



60000769544
Psicologia

PT 540



UMI Number: U607367

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U607367

Published by ProQuest LLC 2014. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346



50000769544

Biblioteca Depòsit

T. 1503

D. 769535
L. 769544

INDICE

CAPÍTULO 1. INTRODUCCIÓN	1
1.1. La medición en Psicología	1
1.1.1. Revisión histórica	1
1.1.2. Sistemas estándar de medición	9
1.1.3. Uso e importancia de los tests en Psicología	16
1.1.4. El proceso de inferencia en los tests	19
1.2. Sistemas de puntuación de las escalas de respuesta	21
CAPÍTULO 2. MÉTODO Y OBJETIVOS	33
2.1. Objetivos	33
2.1.1. La medición en personalidad	35
2.2. Muestra	37
2.3. Variables e instrumentos de medida	38
2.4. Análisis	44
2.5. Diseño	46
CAPÍTULO 3. RESULTADOS: ESTUDIO PRIMERO	53
3.1. Análisis factorial confirmatorio	55
3.2. Análisis con el modelo nominal de Bock (Bock, 1972)	59
3.3. Conclusiones	66
CAPÍTULO 4. RESULTADOS: ESTUDIO SEGUNDO	69
4.1. Análisis factorial confirmatorio	71
4.2. Análisis con los modelos de la teoría de respuesta al ítem	74
4.3. Conclusiones	81
CAPÍTULO 5. RESULTADOS: ESTUDIO DE LAS RAZONES	83
CAPÍTULO 6. DISCUSIÓN	97
REFERENCIAS BIBLIOGRÁFICAS	107



CAPÍTULO 1.

INTRODUCCIÓN

1.1. La medición en Psicología

1.1.1. Revisión histórica

El problema de la medición de variables psicológicas ha estado siempre rodeado de una gran controversia. Es muy frecuente encontrar personas que opinan que la psicología humana es demasiado compleja como para reducirla a números, ya que de este modo se pierde gran parte de la información relativa a la realidad que se pretende medir. Representar la realidad psicológica con números ha sido considerado como un obstáculo para conocer de verdad dicha realidad porque ésta es mucho más rica que el conjunto de los números que la pretenden representar. Posturas como esta han puesto en tela de juicio la utilidad de los métodos cuantitativos en Psicología porque atacan la base de los mismos, es decir, la medición. Por

ello, los psicólogos han tenido que realizar un gran esfuerzo para establecer las bases lógicas que justifican la medición de ciertas variables psicológicas.

Mientras que la medición de algunos atributos del mundo físico, como el tiempo o la distancia, se viene aceptando como válida desde hace siglos, la medición de atributos psicológicos no se comenzó de modo sistemático hasta el siglo XIX, y sólo a mediados del siglo XX ha comenzado a presentar una aceptación amplia en el mundo científico.

Hasta el siglo XX la medición se basaba en el idealismo platónico. Según dicha corriente filosófica, los objetos son poseedores de una cantidad o magnitud de la propiedad que se quiere medir, y para medirla había que asignar a ese objeto un número que representara de la manera más precisa posible la magnitud de dicha propiedad.

En el ámbito de la Psicología, el origen de los intentos de medición de manera sistemática se da en cuatro áreas diferentes (Jáñez, 1989; Santisteban, 1990): la psicofísica, área en la que Fechner aporta un método para la medición de la sensación; la medición del tiempo de reacción en los procesos mentales por parte de Donders; el análisis cuantitativo del aprendizaje, iniciado por Ebbinghaus; y el estudio de las diferencias individuales mediante la utilización de los tests mentales por Galton.

En el ámbito de las ciencias físicas, esta corriente filosófica se plasmó en una serie de axiomas propuestos por Hölder (1901), mediante los cuales se establecían las condiciones que una magnitud debía satisfacer para poder ser medida. Algunos de estos axiomas requieren el establecimiento de

relaciones de orden o igualdad entre cantidades inindividuales, y los axiomas restantes requieren la condición de que se puedan sumar cantidades para obtener una nueva cantidad.

La primera exposición completa de los fundamentos lógicos de la medición física se debe a Campbell (1928), que consideraba la medición como la demostración de que las relaciones identificadas en el sistema numérico. Los requisitos para que esto pudiera llevarse a cabo exigían la posibilidad de establecer relaciones de orden o igualdad entre magnitudes, y también la de aditividad.

Posteriormente, Stevens (1946, 1951) planteó un nuevo concepto de medición según el cual medir no consistía únicamente en buscar un número que representase la cantidad de magnitud presente en un objeto, sino en asignar números a objetos de acuerdo con una serie de reglas, de tal manera que los números asignados en la medición no representan cantidades, sino relaciones entre los objetos. En función de dichas reglas, Stevens extendía el sistema de Campbell permitiendo cuatro escalas de medición, nominales, ordinales, de intervalos y de razón, dependiendo de las operaciones empíricas necesarias para crear cada escala y de las transformaciones posibles dentro de cada tipo de escala. Estas transformaciones debían preservar el isomorfismo entre el sistema empírico y la escala, de tal manera que las relaciones observadas entre los objetos debían verse reflejadas en las relaciones entre los valores numéricos.

La contribución principal de las teorías axiomáticas es que la medición se ve como parte integral de la teoría, más que como un problema

que ha de resolverse antes de la construcción teórica. De este modo, la medición sólo es precisa si lo es la teoría, y no es considerada como un problema que ha de resolverse antes de elaborar la teoría. En este sentido, la medición es un subproducto de la teoría. Solamente cuando los datos satisfacen los supuestos de la teoría se puede obtener la medición (Coombs et al., 1981).

De este modo la medición se considera como la construcción de un modelo formal de alguna característica o propiedad del mundo real, con la que se pretende representar las relaciones existentes en dicho sistema empírico (Muñiz, 1992). Si se han empleado los números como sistema formal, entonces la representación se denomina medición.

Es decir, los constructos teóricos (como la ansiedad, el conservadurismo, la inteligencia, el neuroticismo, etc.) están conectados entre sí mediante algún modelo formal que les da una estructura, pero al mismo tiempo están conectados con los fenómenos directamente observables. Según Lord y Novick (1968), los constructos se relacionan entre sí mediante definiciones sintácticas, y con los fenómenos observables por medio de definiciones semánticas. Antes de medir hay que identificar los elementos del mundo real con los constructos del sistema teórico por medio de una definición semántica precisa, y a continuación se puede proceder a la medición.

Para estos autores, la medida es “un procedimiento para la asignación de números (puntuaciones, medidas) a propiedades especificadas de unidades experimentales, de modo que las caractericen y preserven las

relaciones especificadas en el dominio comportamental” (Lord y Novick, 1968, pág. 17). Para Stevens (1951), la medida consiste en asignar números a objetos o sucesos en función de ciertas reglas. Estas reglas, o lo que Lord y Novick llaman preservar las relaciones, suponen que para representar dicha propiedad debe existir un isomorfismo entre las características del sistema numérico y las relaciones entre las distintas cantidades o magnitudes de la propiedad medida.

En este sentido, Suppes y Zinnes (1963) presentan una definición de la medición según la cual este proceso consiste en representar un sistema relacional empírico mediante un sistema relacional numérico. El sistema relacional empírico hace referencia a un conjunto de magnitudes del atributo o propiedad que se pretende medir, y el sistema relacional numérico designa el grupo de números que se utiliza para representar las magnitudes. La medición consistiría en convertir la información referida al sistema relacional empírico en información expresada mediante un sistema numérico.

Estas definiciones son similares porque todas hacen referencia a un atributo que hay que medir, a los números que resultan de dicha medición y a la noción de que las magnitudes correspondientes al atributo en cuestión son reflejadas en un grupo de números.

Este enfoque de este tipo de teorías de la medición es representacional en el sentido de que los símbolos asignados a los objetos deben representar las relaciones observadas entre las propiedades de los objetos, ya que lo que se pretende es que dicha asignación constituya una

medición. Esta moderna perspectiva de la teoría de la medición (Teoría Representacional de la Medición) trata el tema de la medida abordando tres áreas: la representación, la unicidad y la significación.

Desde el punto de vista de la representación, la medición supone encontrar un sistema relacional numérico con una estructura semejante a la del sistema relacional empírico que se pretende medir. Debido a esta semejanza, el primer sistema puede representar al segundo. La medición consistirá en encontrar un sistema relacional numérico que pueda representar al sistema relacional empírico que se pretende medir (el que más se asemeje al sistema empírico). O lo que es lo mismo, la representación numérica consistirá en la asignación de números a objetos de tal modo que las relaciones observadas entre los objetos sean reflejadas por las relaciones existentes entre los números que se les han asignado (Muñiz, 1992).

Así pues, cuando se asignan números a una realidad psicológica no se está pretendiendo reducir esa realidad a números, sino simbolizar numéricamente algunos aspectos de dicha realidad y algunas de todas las posibles relaciones existentes entre sus componentes. Es decir, se pretende codificar en lenguaje numérico una cierta cantidad de información sobre la realidad que se está estudiando, ya que es difícil captar por completo toda la información de una realidad psicológica. Y aunque no se puedan medir todos los aspectos de la realidad, a través de la medición se puede obtener más información acerca de dicha realidad (Jáñez, 1989).

En segundo lugar, la Teoría Representacional estudia el problema de la unicidad, que hace referencia a la posibilidad de que existan distintas

escalas numéricas para integrar el sistema relacional numérico. Es decir, una vez establecidas las relaciones numéricas que pueden representar a una realidad empírica, se pueden obtener distintas escalas de números que la representen igualmente. La cuestión está en cómo puede cambiarse un conjunto de números por otro sin que resulten modificadas las relaciones numéricas que representan las relaciones empíricas.

Las transformaciones admisibles de las relaciones existentes entre los números utilizados para representar la realidad empírica permiten clasificar las escalas de medida a que dan lugar dichas relaciones. En función del tipo de escala se podrán llevar a cabo ciertos análisis matemáticos con los números que representan dicha realidad, y también dependerá del tipo de escala la interpretación de los resultados obtenidos a partir de los análisis realizados. Así pues, en función del tipo de transformación u operación empírica que se puede realizar con los datos medidos se pueden considerar cuatro escalas principales:

- *Nominal*, que permite la determinación de una relación de igualdad y desigualdad, y admite cambiar los números asignados a los objetos por otro conjunto de números que preserve dicha relación. Por tanto, dados dos elementos a y b , se podrán considerar equivalentes si y sólo si $f(a) = f(b)$, es decir, el conjunto de transformaciones admisibles será el de las aplicaciones que son inyectivas.

- *Ordinal*, que además permite determinar relaciones de orden y admite transformaciones que preserven dicho orden. Por tanto, dados dos elementos a y b , se podrán considerar equivalentes si y solo si $f(a) = f(b)$,

pero además se podrá considerar que $a > b$ si y sólo si $f(a) > f(b)$, es decir, el conjunto de transformaciones admisibles es el de las aplicaciones que sean crecientes.

- *De intervalo*, que además permite determinar la igualdad de las diferencias o intervalos existentes entre las puntuaciones de la escala y admite la realización de transformaciones lineales positivas, ya que estas dejan invariante la igualdad de los intervalos. Es decir, el conjunto de transformaciones admisibles es el de las aplicaciones de la forma $f(x) = ax + b$, con $a > 0$.

- *De razón*, que permite determinar la igualdad de razones y admite la multiplicación de los números por una constante positiva, lo que preserva la igualdad de la relación entre el sistema numérico y el empírico. Además asume la existencia de un cero absoluto que indica ausencia de la propiedad o característica que se está midiendo. Es decir, el conjunto de transformaciones posibles es el de las aplicaciones de la forma $f(x) = ax$, con $a > 0$.

El tercer aspecto del que se ocupa la Teoría Representacional es el de la significación, que hace referencia a la validez de una conclusión empírica que es inferida a partir de una conclusión numérica. Para hacer cualquier inferencia a partir de los datos numéricos es preciso conocer la escala que los ha originado, ya que la validez de cualquier conclusión dependerá del tipo de escala. Por ejemplo, no se pueden comparar dos medias a menos que las propiedades a las que representan estén medidas en una escala de intervalo, ya que de lo contrario las conclusiones a las que se

llegara no serían invariantes respecto a las transformaciones admisibles de la escala.

Por su parte, Lord y Novick (1968) realizan una propuesta más práctica e insisten en la necesidad de evaluar los datos desde el punto de vista de la utilidad de la escala resultante, señalando que si las puntuaciones totales de los tests proporcionan más información para la clasificación de los sujetos y la predicción cuando son tratadas como una escala de intervalo, se pueden emplear así. Si una escala de intervalo permite realizar predicciones precisas y una descripción útil del modelo formal empleado para representar la realidad, entonces es una buena escala y futuros desarrollos teóricos deberían basarse en ella.

En cualquier caso, esta aproximación a la medida centrada en el sujeto (Torgerson, 1958) tiene como objetivo fundamental situar a los sujetos en algún punto de un continuo que representaría al rasgo o atributo psicológico no observable, además de diferenciar a los sujetos. Para ello las puntuaciones de los tests deberán presentar la mayor fiabilidad posible (menor cantidad de errores), además de mantener los requisitos de orden y aditividad en la variable medida, lo que permitiría situar a los sujetos en el continuo del rasgo y diferenciarlos.

1.1.2. Sistemas estándar de medición

Las teorías de la medición intentan explicar cómo y bajo qué circunstancias es posible emplear números para representar la información

concerniente acerca de una serie de atributos o propiedades. Estas teorías pueden clasificarse en función del grado de generalidad de las mismas. Teorías como la de Campbell (1928), la de Stevens (1951) y la de Suppes y Zinnes (1963), mencionadas anteriormente, se refieren a la medición en general. Sin embargo otras teorías se enmarcan en áreas especializadas o en ciertos aspectos de la medición, como por ejemplo la de Lord (1953), la de Torgerson (1958) o la de Wherry (1982), que hacen referencia, respectivamente, a la teoría del rasgo latente en los tests, a los métodos de escalamiento y a la medición en personalidad mediante escalas de valoración.

Como se ha indicado anteriormente, Campbell (1928) estipuló que las relaciones de orden y aditividad se pueden determinar y comprobar empíricamente entre magnitudes físicas. Sin embargo, en el caso de otras propiedades esto no es posible hacerlo directamente, por lo que es necesario confiar en mecanismos para identificar el orden, la igualdad o la ausencia de una magnitud. Estos mecanismos son sistemas estandarizados de medición, y se emplean para medir una propiedad o atributo de interés.

Según Aftanas (1988), un sistema estándar de medición sería “cualquier artificio, mecanismo o proceso discriminativo que puede ser utilizado para denotar e indicar la extensión de magnitud, o las diferencias entre magnitudes, de una propiedad”. Hablar de proceso discriminativo hace referencia a la posibilidad de que el juicio humano pueda ser utilizado como un sistema estándar básico para estimar las magnitudes o diferencias entre magnitudes. Con el término denotar se hace referencia al acto de encontrar

las magnitudes de la propiedad mediante el sistema estándar, es decir, el sistema estándar debe encontrar o entrar en contacto con dicha propiedad.

El término “sistema estándar” hace referencia a un mecanismo de medición generalizado, no a un mecanismo asociado a un área especializada. El término “sistema” se emplea en el mismo sentido que lo hacían Suppes y Zinnes (1963), haciendo referencia a los sistemas relacionales empírico y numérico. Un sistema estándar es un mecanismo o artificio que permite convertir la información acerca de un sistema empírico en información representada mediante un sistema numérico. El término “estándar” se utiliza en el sentido de una unidad común de medida. Así pues, un sistema estándar es un mecanismo comúnmente aceptado para la medición de un atributo.

Cuando se emplea cualquier sistema estándar para establecer el orden o la igualdad entre magnitudes, está justificado establecer un enunciado numérico acerca de dichas magnitudes. Campbell (1928) proponía la observación humana como sistema estándar para identificar la igualdad de dos magnitudes, por ejemplo de la longitud de dos carreteras.

Sin embargo, otros sistemas estándar pueden ser utilizados para identificar el orden o la igualdad entre magnitudes, además de la observación humana. Cuando dicho sistema puede identificar este tipo de relaciones entre propiedades, se puede formular un enunciado numérico acerca de las mismas. En el caso de la observación humana, la evaluación de los enunciados formulados implicaría la confirmación, por parte de otros observadores, de las relaciones encontradas. Por ejemplo, en la observación de longitudes, si varias personas están de acuerdo en que $X > Y$ y ninguna

discrepa, el investigador puede considerar razonable el asignar un número mayor a X.

Sin embargo, en el caso de que el sistema estándar no sea la observación directa del atributo, para verificar los enunciados numéricos formulados hay que comprobar que otros sistemas estándar que pretenden medir dicho atributo llegan a la misma conclusión y, si esto ocurre, hay que comprobar si el sistema estándar es un buen indicador del atributo mediante un proceso de validación.

En primer lugar, el proceso de medición requiere la *identificación de un sistema estándar de medición apropiado*. En el caso de los atributos psicológicos, el proceso de medición incluye el desarrollo o utilización de sistemas estándar que pueden ser de distintos tipos pero que tienen la misma complejidad, ya que la medición de atributos psicológicos es compleja en cualquier caso.

En segundo lugar, el proceso de medición requiere *que el sistema estándar sea capaz de identificar la presencia o ausencia del atributo de interés*, el cual no es directamente observable. Es decir, la propiedad o atributo ha de poder ser puesta de manifiesto mediante un sistema estándar porque no es directamente observable. Podría hablarse de la medición en términos de la interacción que tiene lugar entre un sistema estándar y el atributo que éste pretende medir. En psicología, la mayoría de las medidas incluyen la interacción entre el sistema estándar y el atributo. Por ejemplo, en los tests psicológicos es precisamente la interacción entre el test y la

persona la que hace posible la medición de atributos latentes, que no son directamente observables.

Además, debe haber un *acuerdo científico* respecto a la posibilidad de que el atributo se ponga de manifiesto mediante un sistema estándar particular, y respecto a la precisión de la medición realizada. Puede existir desde un completo acuerdo hasta un completo desacuerdo. En los casos en los que la evidencia es clara, como cuando se mide una longitud mediante una regla, el acuerdo puede ser muy alto, ya que existe un acuerdo generalizado en torno a lo que es la propiedad de la longitud y en torno a la adecuación de medirla mediante una regla. Sin embargo, cuando se utilizan sistemas estándar como el autoinforme para medir características de personalidad el acuerdo tiende a ser bastante bajo.

En tercer lugar, se debe *especificar un modelo* para convertir la información obtenida a partir de la medición que provee el sistema estándar en información numérica. Cuando un sistema estándar ha sido elaborado en torno a una unidad común de medida la información métrica se obtiene directamente a partir de la utilización de dicho sistema, como ocurre por ejemplo con un test de habilidades que contenga items calibrados. En otras ocasiones la información métrica ha de deducirse indirectamente mediante un modelo que permita la conversión de la información ofrecida por el sistema estándar. La medición obtenida por el sistema estándar sirve como base para la representación de las magnitudes mediante el sistema numérico, y las características específicas del sistema numérico vienen dadas por el tipo de información ofrecida por el sistema estándar. Stevens (1946) y

Torgerson (1953), entre otros, propusieron varias de las representaciones métricas básicas.

En cuarto lugar, deben desarrollarse *criterios* y procedimientos para determinar en qué grado el resultado de la medición refleja y representa con precisión el atributo que se ha medido. Estos criterios evalúan si los resultados ofrecidos por el sistema estándar son verificables o replicables a través de sistemas estándar independientes, evaluando empíricamente si las diferentes cantidades indicadas por el sistema estándar reflejan con precisión diferentes magnitudes del atributo de interés.

Respecto a la medición en áreas especializadas de la psicología, se puede distinguir entre tres tipos de sistemas estándar: el elemental, el planeado y el de proceso dual. El primer tipo es la propia capacidad humana para distinguir entre magnitudes de atributos, mientras que el segundo tipo está constituido por instrumentos contruidos por los psicólogos para poner de manifiesto magnitudes de ciertas características de los individuos, entre los que se incluirían los tests de habilidades, los de rendimiento y muchos instrumentos físicos que se emplean en psicología para medir atributos comportamentales. Éstos últimos sirven como criterio para contrastar los resultados obtenidos a partir de los tests mencionados.

El tercer tipo está asociado con los instrumentos de autoinforme, que incluyen dos procesos o sistemas estándar diferentes. Por una parte, incluye un sistema estándar que consiste en una serie de cuestiones desarrolladas por los psicólogos, y que representaría un sistema estándar independiente. Por otra parte, la interacción entre el sujeto que responde y el sistema

estándar, lo que requiere una automedición que puede ser considerada como un sistema estándar de medición aparte. El proceso dual de la medición de las características de personalidad se puede distinguir del proceso dual de medición de preferencias e intereses en que el primero mide el comportamiento típico de un individuo, mientras que el segundo se centra en la evaluación que una persona hace de un objeto o situación.

En el proceso dual de medición de la personalidad, el prototipo de esta situación de medición es el cuestionario o test de autoinforme de personalidad. El sistema estándar consiste en un conjunto de cuestiones o enunciados con los cuales el individuo que responde debe mostrar su acuerdo en un formato de verdadero-falso, o indicar su cantidad de acuerdo en un formato de respuesta tipo Likert. Desde la perspectiva del análisis de un sistema estándar, esta situación de medición incluye dos tipos de procedimientos de estimación. El primero está asociado al sistema estándar independiente constituido por el grupo de items y el segundo está asociado con las respuestas a cada uno de los items. La magnitud final de interés directo para el psicólogo es la cantidad en que los sujetos que responden al test poseen la característica medida.

Un modelo de este tipo debe considerar tanto las características del sistema estándar constituido por los items como las respuestas dadas en la autoevaluación. Para convertir o transformar las evaluaciones o resultados obtenidos mediante un sistema estándar en información métrica o numérica han sido elaborados distintos modelos. En otros casos, sin embargo, no se ha desarrollado completamente un modelo para algunas situaciones de medición. Por ejemplo, en las escalas de valoración se asume que la

información métrica o numérica se deriva directamente a partir de las valoraciones obtenidas mediante el sistema estándar, es decir, a partir del valor que escoge el sujeto en una escala de cinco o siete puntos. Un ejemplo de un modelo completamente desarrollado sería la Ley del Juicio Comparativo de Thurstone (1927), que permite la transformación de un juicio comparativo en una métrica de intervalos iguales, aunque sería un modelo aplicable a un sistema estándar elemental.

Por otra parte, para establecer la precisión con que se realice la medición habría que establecer si el resultado de dicha medición incluye error. La existencia de error puede ser indicada cuando distintos sistemas estándar presentan desacuerdo en las magnitudes obtenidas o cuando la relación entre dichas magnitudes y los indicadores de dichas magnitudes ofrecidos por dicho sistema no se pueden establecer de un modo claro. Wiggins (1973), examinando el grado de error de los tests de personalidad de tipo autoinforme, identificó dos componentes de variación en la medición: las características del ítem, como la deseabilidad social del ítem y la asignación de valores a las categorías de la escala de respuesta al ítem, y los estilos de respuesta o tendencias individuales a dar un tipo de respuesta determinada.

1.1.3. Uso e importancia de los tests en Psicología

Los tests y demás instrumentos de medición forman parte del quehacer cotidiano del psicólogo, ya que se emplean con asiduidad en ámbitos como la educación, la selección de personal, la orientación

profesional, etc. Los resultados obtenidos a partir de dichas mediciones suelen tener bastante repercusión en la vida de las personas. Por ello el psicólogo debe conocer los procedimientos en los que se basan sus puntuaciones, la información que proporcionan y sus limitaciones. Si se utilizan incorrectamente, los tests pueden ser instrumentos que lleven a un psicólogo a un grave error a la hora de la toma de decisiones.

Aunque en todas las ciencias el conocimiento de los instrumentos de medida es fundamental, el psicólogo debe hacer un esfuerzo mayor en este sentido debido a que los atributos psicológicos no pueden medirse directamente, y sin embargo se utilizan para explicar la conducta humana. El grado en que un individuo está caracterizado por estos constructos debe inferirse a partir de la observación de su conducta.

Para ello, antes de medir el constructo en cuestión hay que establecer una correspondencia entre dicho constructo y las conductas observables que constituyen un indicador del mismo (Lord y Novick, 1968), es decir, hay que definir operacionalmente el constructo. A continuación hay que obtener una muestra de las conductas consideradas indicadores del constructo en situaciones estandarizadas y asignar un valor cuantitativo a la muestra de conductas recogidas mediante el test. A partir de este valor, que constituye una medida del atributo psicológico, el psicólogo realiza una inferencia acerca de la cantidad que caracteriza al sujeto y que le sitúa en un punto en el continuo del constructo teórico.

Puesto que los constructos psicológicos son conceptos teóricos que únicamente se pueden evaluar indirectamente, a la hora de elaborar

instrumentos para la medida de constructos psicológicos hay que tener en cuenta una serie de cuestiones importantes. Por una parte, no hay una única aproximación a la medida de un constructo que se acepte universalmente, ya que se pueden seleccionar diferentes conductas como indicadores del mismo constructo y se pueden emplear distintos procedimientos de medida en función de dichos indicadores, con lo que se pueden obtener conclusiones diferentes a la hora de situar a un sujeto en el continuo del rasgo psicológico que se pretende medir. Esto implica, además, que las medidas psicológicas están basadas en muestras limitadas de conducta y no en todas las conductas posibles, por lo que es importante elegir una muestra que sea representativa de dichas conductas.

Por otra parte, la medida que se obtiene siempre tiene una parte de error, precisamente porque el número de conductas que se utilizan para medir el constructo es limitado, ya que no se incluyen todas las conductas posibles, y porque la medida se realiza en un momento concreto del tiempo. Por ello es importante poder estimar el grado de error que se halla en un conjunto de observaciones.

Además, los constructos teóricos no tienen sentido si se definen aisladamente. Deben formar parte de una teoría psicológica, que es un sistema de proposiciones referidas a un conjunto de constructos teóricos conectados entre sí mediante un modelo formal, y que sirven para explicar algunos aspectos de la realidad, por lo que también están conectados con fenómenos directamente observables. Hay que definir claramente las propiedades de la escala de medida para poder interpretar adecuadamente los resultados obtenidos, y hay que poner en relación el constructo que se

está midiendo con otros constructos teóricos y con los fenómenos observables.

En función de todo lo anterior, podría decirse que los tests pretenden ser medidas objetivas y precisas de algún constructo teórico. Dichas medidas se obtienen a partir de una muestra representativa de conductas consideradas como indicadores de dicho constructo en situaciones estandarizadas. Son además una técnica que permite recoger la información de forma sistemática, siguiendo siempre las mismas instrucciones y respondiendo los sujetos a las mismas cuestiones, evitando así que las diferencias entre las puntuaciones obtenidas se deban a diferencias en la aplicación del test. La puntuación obtenida por los sujetos en el test se estima comparándola con la puntuación que presenta un grupo normativo en el rasgo medido o con la puntuación en algún criterio externo, implicando normalmente la realización de predicciones o inferencias acerca de otras conductas

1.1.4. El proceso de inferencia en los tests

El objetivo del proceso de medida en Psicología es describir una característica o atributo de un sujeto mediante una puntuación numérica que representa la cantidad o magnitud de dicho atributo que posee el sujeto. La representación cuantitativa de estas características permite establecer comparaciones entre sujetos y distintos análisis matemáticos empleando técnicas estadísticas. Cuando una conducta es abierta se puede observar directamente, y se registra asignándole una puntuación en una escala; si la

conducta es encubierta, es decir, no es directamente observable, se administra un test o un cuestionario y se obtiene así una serie de respuestas que se transforman en una puntuación numérica.

Al obtener de este modo una puntuación del sujeto en el rasgo, se está asumiendo que dicha puntuación refleja la cantidad de interés que el sujeto posee en la característica o atributo medido (Martínez Arias, 1995). Sin embargo, no se sabe realmente si dicha puntuación numérica refleja bien la verdadera cantidad de interés, ya que siempre está presente la influencia del azar en forma de errores aleatorios. Por esta razón, el proceso de medición en Psicología es en realidad un proceso de inferencia.

Dicho proceso parte, en primer lugar, de la construcción de las cuestiones que van a formar parte del test, y a las cuales ha de responder el sujeto. Las cuestiones que se formulan y las respuestas del sujeto constituyen la única parte directamente observable del proceso de medición, y a partir de estas respuestas se describe el rasgo que se está midiendo otorgándole una puntuación al sujeto. Dicha puntuación es considerada una puntuación observada, ya que se obtiene en la parte del proceso de medición que es directamente observable y a partir de una muestra de cuestiones de todas las que sería posible formular. Por tanto, la puntuación observada, que puede representar mejor o peor el valor auténtico en el rasgo que presenta el sujeto, se considera un estimador muestral de dicho valor. Si se pudiesen formular al sujeto todas las cuestiones posibles se podría obtener una puntuación verdadera en el rasgo, pero como no es posible hacerlo se asume que la puntuación observada es un buen indicador de la puntuación verdadera.

El grado en que la puntuación observada representa a la puntuación verdadera en el rasgo es lo que se llama fiabilidad de un test, y la validez indica la capacidad de la puntuación verdadera para reflejar lo que el test intenta medir, que es un constructo teórico, es decir, un rasgo del sujeto, el cual normalmente no es observable (Martínez Arias, 1995). En la medida que la puntuación observada no represente fielmente a la puntuación verdadera del sujeto en el rasgo, se verán afectadas igualmente la fiabilidad y la validez del instrumento de medida empleado.

1.2. Sistemas de puntuación de las escalas de respuesta

Según la Teoría Representacional de la Medición, cuando se está procediendo a medir algún constructo teórico las relaciones que presentan los elementos empíricos han de verse reflejadas en el sistema numérico utilizado para representarlas. En función del tipo de relaciones que se representen, se podrá realizar una serie de transformaciones matemáticas de los datos, dependiendo de las cuales se puede hablar de escalas nominales, ordinales, de intervalos y de razón.

Como se ha indicado anteriormente, si se pretende medir una conducta que no es directamente observable, generalmente se administra un test o un cuestionario. A partir de las respuestas se obtiene una puntuación numérica del sujeto en el rasgo, asumiéndose que dicha puntuación refleja el auténtico valor que presenta el sujeto en dicho rasgo. Este proceso de

transformar las respuestas en puntuaciones se conoce como escalamiento, y su objetivo principal es producir buenas escalas (Allen y Yen, 1979). El número asignado a un objeto particular o a un sujeto se denomina valor de escala. Dichos valores son los que han de preservar las relaciones del sistema empírico al que representan, y cuando no lo hacen vulneran el isomorfismo que ha de existir entre los sistemas empírico y numérico, afectando a la fiabilidad y a la validez del test.

En este sentido, se ha estudiado poco el sistema de puntuación de diversos tipos de escalas de respuesta ordinales, y en concreto, la puntuación y significado que se le atribuye al punto medio o punto neutro que se emplea en las escalas de respuesta de valoración (“rating scales”). Este tipo de escalas se utiliza para medir un amplio rango de actitudes y de características de personalidad, e incluyen escalas de diversos tipos. Un tipo de escala de valoración que se ha utilizado ampliamente tanto en contextos de investigación como aplicados es la escala de actitud (Aiken, 1996). Este tipo de escala consiste en un grupo de enunciados positivos y negativos relativos a una materia de interés y que se responde sobre un continuo consistente en un número de categorías. La más común de todas las técnicas de medida de actitud es la escala Likert.

El requisito fundamental cuando se utiliza una escala de tipo Likert es que las categorías de respuesta estén ordenadas e igualmente espaciadas a lo largo de un continuo unidimensional. Estas categorías han de ser aplicables a todos los items del instrumento de medida de la actitud, de manera que se puedan asignar valores enteros consecutivos a dichas categorías de respuesta ordenadas.

Aunque Likert (1932) no basó su procedimiento de escalamiento en un modelo de respuesta explícito, demostró empíricamente que era adecuado asignar valores enteros a las categorías de respuesta ordenadas. Andrich (1982) observó que aunque la relación entre la puntuación total en la escala de actitud y el nivel de rasgo estimado mediante un modelo Rasch no es lineal, hay un amplio espectro del nivel del rasgo dentro del cual la relación es lineal. Esto explicaría que se pudieran asignar valores enteros a las categorías y que la puntuación total en la escala se aproximara a una escala de medida de intervalo.

En una escala Likert, el formato de las respuestas a un ítem tradicionalmente consiste en uno o más niveles de acuerdo, uno o más niveles de desacuerdo y un nivel de indecisión o indeterminación, representado por un interrogante (?) o por categorías de respuesta como “indeciso” o “neutral”. En la literatura se ha estudiado ampliamente las propiedades de una escala completa o las de cada uno de los ítems, pero se ha estudiado poco el significado que tiene para los sujetos cada categoría de la escala de respuesta. Normalmente se asignan valores ordenados a estas categorías de respuesta asumiendo que se trata de una escala de intervalo y que el sujeto que responde atribuye a cada categoría de respuesta el mismo significado atribuido por el investigador. Sin embargo, algunas investigaciones sugieren que asumir este tipo de atribuciones puede no ser válido (Wang y Stanley, 1970). Es necesario considerar no sólo la puntuación obtenida en cada ítem, sino también considerar cómo interpreta el sujeto cada ítem y sus alternativas de respuesta (Dubois y Burns, 1975).

El problema del significado y de la puntuación de la categoría de respuesta “?” sería una cuestión a tener en cuenta en este sentido. De acuerdo con la poca investigación llevada a cabo acerca de este tema (Cronbach, 1946; Edwards, 1946; Worthy, 1969; Goldberg, 1971; Kaplan, 1972) el significado esencial de la categoría “?” sería el de ambivalencia o el de indiferencia. El sujeto ambivalente es el que presenta sentimientos positivos y negativos acerca de la misma cuestión, con lo que la ambivalencia sería el resultado de un alto nivel de implicación con el objeto, mientras que el sujeto indiferente elige la interrogación porque no se siente interesado por la cuestión.

Sin embargo, podría haber otras razones para elegir esta categoría de respuesta (Dubois y Burns, 1975). Por ejemplo, algunas personas pueden elegir la “?” porque no se sienten suficientemente competentes o suficientemente informados como para elegir alguna de las otras dos categorías de respuesta. Otros pueden elegirla para indicar su rechazo a revelar cuestiones personales, y otros pueden elegirla porque no entienden la cuestión que se les plantea.

En este sentido, Cruickshank (1978) encontró evidencias de que la elección de la interrogación como respuesta en una lista de adjetivos sobre el estado de ánimo podía ser debida al bajo nivel de vocabulario de los pacientes, de modo que aquellos que no entendían bien el significado de los adjetivos que se les presentaban escogían la interrogación como respuesta. Con el objeto de estudiar los cambios en el estado de ánimo en una muestra de pacientes, este investigador administró a los pacientes esta lista de

adjetivos antes y después del tratamiento. De esta lista de adjetivos se derivan sendas escalas de estrés y activación que contienen un número desigual de adjetivos positivos y negativos, con lo que puede existir un sesgo. Además, las instrucciones acerca del sistema de puntuación podían dar lugar a más sesgo ya que las respuestas “++” y “+” eran tratadas como positivas y las respuestas “?” y “-” como negativas. En el análisis factorial realizado encontró cuatro factores: alto nivel de estrés, bajo nivel de estrés, alto nivel de activación y bajo nivel de activación.

Cruickshank decidió estudiar la incidencia de la “?” y observó que la frecuencia de una palabra, utilizada como estimador de la familiaridad de la misma, presentaba una correlación significativa con el uso de la interrogación como respuesta. Al eliminar las interrogaciones del análisis, considerándolas como datos faltantes, la distinción que previamente encontró entre estrés y nivel de activación desapareció. Tras eliminar los items que presentaban más de un 15% de interrogantes como respuesta realizó de nuevo los análisis, y obtuvo de nuevo los factores de estrés y nivel de activación. A partir de estos resultados dedujo que cuando hay una gran proporción de “?”, el modo en que se puntúan es crucial para la interpretación de los resultados.

Así pues, no está claro el significado que se le atribuye a la interrogación, y además su utilización masiva como respuesta a un grupo de items puede alterar los resultados de los análisis que se vayan a llevar a cabo. Dubois y Burns (1975) realizaron un estudio para comprobar si el significado que los sujetos atribuían a la “?” era realmente el de punto intermedio. Si esto era así, estos autores consideraban que sería razonable

esperar que la variabilidad de las puntuaciones totales obtenidas para aquellos sujetos que eligieron como respuesta la “?” debería ser similar a la variabilidad de las puntuaciones totales obtenidas por los sujetos que eligen cualquier otra categoría de respuesta. Por otra parte, también consideraron que sería razonable esperar que la media de la distribución de las puntuaciones totales para aquellos sujetos que responden con la “?” debería ajustarse al patrón de medias sugerido por la distribución de las puntuaciones totales correspondientes a los sujetos que responden a otras categorías.

Los resultados mostraron que había diferencias estadísticamente significativas entre la variabilidad de la “?” y la variabilidad de otras categorías en cuatro de las once escalas que utilizaron en la investigación. Además, las medias de las puntuaciones totales obtenidas para cada categoría de respuesta se distribuían a lo largo de una línea recta en la cual la media correspondiente a la categoría “?” se situaba en el medio. Estos resultados indicaban que los sujetos que elegían la categoría “?” como respuesta se situaban en el punto intermedio de la escala, con lo que se daba soporte empírico al sistema de puntuación habitual de la “?” considerándola como punto medio de la escala.

Sin embargo, aunque la categoría “?” pueda ser considerada como una categoría intermedia, eso no significa que sea únicamente un indicador de dicha posición, como asumen muchos investigadores, ya que también encontraron evidencias de la presencia de un estilo de respuesta “?” y de respuestas que indicarían ambivalencia o indiferencia.

Según algunos investigadores, la inclusión de la categoría de respuesta “no estoy seguro/indeciso” en mitad de la escala de respuesta representando un punto neutral amenaza el requisito de la unidimensionalidad (Cheung y Mooi, 1994), lo cual puede ser debido a la posible existencia de un estilo de respuesta caracterizado por la elección sistemática del punto medio de la escala de respuesta.

Por ello, a la hora de asignar un valor a la interrogación hay diversas posturas. Cronbach (1946), por ejemplo, ha sugerido que dada la ambigüedad del significado que puede presentar esta categoría, no debe seguir utilizándose. Meddis (1972) señala que la interrogación es difícil de clasificar y que lo mejor sería tratarla como dato faltante en listas de adjetivos, ya que su elección como respuesta se debe a una falta de entendimiento del ítem planteado. En una investigación realizada, Goldberg (1971) preguntaba a los sujetos las razones por las cuales habían escogido la interrogación, en el caso de que lo hubieran hecho, ofreciendo una lista de posibles razones para que escogieran la más adecuada en cada caso. Sin embargo, otros investigadores, aunque reconocen que los sujetos pueden escoger la interrogación como respuesta por diversas razones, la codifican como si fuera un indicador de la posición intermedia en un continuo.

De hecho, en la mayoría de las investigaciones sobre actitudes se asume o bien que los sujetos eligen la interrogación como indicador de una posición intermedia, o bien se considera que, aunque pueden elegir esta categoría de respuesta por distintas razones, éstas se contrabalancean entre sí de manera que al final el efecto es similar al obtenido si los sujetos

hubiesen utilizado la interrogación con el propósito de indicar su neutralidad.

También en otras escalas, como en el Job Descriptive Index (JDI; Smith, Kendall y Hulin, 1969), figura la interrogación como punto medio en la escala de respuesta. En este caso, el sistema de puntuación que originalmente se empleó en esta escala consistía en dar una puntuación de 3 si se manifiesta acuerdo ante items positivos y desacuerdo ante items negativos, dar una puntuación de 1 si se manifiesta desacuerdo en items positivos y acuerdo en items negativos, y dar una puntuación de 2 a la interrogación. Se consideraba que esta categoría de respuesta era un indicador de la posición intermedia en el continuo.

Sin embargo, este sistema de puntuación (3, 2, 1) fue sustituido por los propios autores de la escala tras comprobar que los sujetos que presentaban baja satisfacción respondían la interrogación un número de veces significativamente mayor que el grupo de sujetos con alta satisfacción, por lo que dicha categoría era más indicativa de insatisfacción que de satisfacción. Por esta razón, el sistema de puntuación pasó a ser 3, 1, 0.

Hanisch (1992) se propuso evaluar, mediante un modelo politómico de la teoría de respuesta al ítem, si el sistema original de puntuación (3, 2, 1) era el apropiado. Por una parte calculó las correlaciones entre el número de respuestas para la muestra total, obteniendo los siguientes resultados: correlaciones negativas significativas entre el número de respuestas negativas y positivas; correlaciones negativas significativas entre el número

de respuestas positivas y el número de “?”, y; correlaciones positivas significativas entre el número de respuestas negativas y el número de “?”.

A continuación, dividió la muestra en dos grupos, satisfechos e insatisfechos, en función de la puntuación media en satisfacción obtenida por los sujetos, y calculó el promedio de respuestas positivas, el promedio de “?” y el promedio de respuestas negativas dadas por los sujetos satisfechos y por los sujetos insatisfechos. La media de “?” en el grupo de sujetos satisfechos fue de 0,88 respuestas, y en el grupo de insatisfechos fue de 3,17 respuestas. La diferencia entre estas dos medias resultó ser estadísticamente significativa ($p < 0,05$), lo que confirmaba los resultados obtenidos previamente por Smith et al. (1969).

Pero además de replicar los análisis de Smith et al., Hanisch (1992) estudió el sistema de puntuación del JDI aplicando un modelo de teoría de respuesta al ítem, concretamente el modelo nominal de Bock (Bock, 1972). Los resultados mostraron que, para 13 de los 18 ítems de la escala, los sujetos que presentaban baja satisfacción tenían más probabilidad de seleccionar la interrogación como respuesta a dichos ítems, que los sujetos con alta satisfacción. Estos resultados apoyaban el hecho de que asumir un sistema de puntuación 3, 2, 1 para las categorías de respuesta de los ítems del JDI no parecía el más apropiado, mientras que el sistema de puntuación 3, 1, 0, propuesto también por los autores de la escala, reflejaba mejor el significado de las diferentes opciones de respuesta y, consecuentemente, de la validez del cuestionario.

Pero además de estudiar el significado atribuible al punto medio en una escala de respuesta, y la puntuación más adecuada para dicho punto medio, otros investigadores han estudiado si las categorías de la escala de respuesta se presentan ordenadas a lo largo de un continuo, tal y como se asume. En este sentido, Andrich y Schoubroeck (1989) estudiaron las características del Cuestionario de Salud General (GHQ; Goldberg, 1978) aplicando un modelo de rasgo latente que permite la estimación de los umbrales existentes entre categorías de respuesta ordenadas, además de comprobar si era adecuado un sistema de puntuación mediante números enteros en una secuencia ordenada.

Los items del GHQ presentan cuatro escalas de respuesta tipo Likert diferentes, con cuatro categorías de respuesta cada una. A dichas categorías se les asignan números enteros en una secuencia ordenada de manera que, a mayor puntuación, menor nivel de salud. Según Bock (1975), sólo cuando las distancias entre los umbrales son iguales está justificado un sistema de puntuación mediante números enteros, con lo que el uso de un sistema de puntuación de este tipo, sin comprobar previamente las distancias entre los umbrales, puede llevar a resultados engañosos.

Para comprobar si realmente las categorías de respuesta se presentaban ordenadas, se estimaron los umbrales existentes entre categorías sucesivas dentro de cada ítem mediante un modelo de teoría de respuesta al ítem desarrollado por Andrich (1978; 1982). En este modelo no se asume un orden para los umbrales, pero si las categorías de respuesta están realmente ordenadas, como se asume en un principio, los umbrales también se presentarán ordenados. La estimación de la distancia entre dos umbrales

consecutivos daría la ponderación apropiada para la correspondiente categoría a la hora de asignarle una puntuación.

Los resultados obtenidos mostraron que los umbrales se presentaban en el orden correcto, lo que confirmaba el orden operacionalizado en las categorías de respuesta. Sin embargo, estos umbrales variaban bastante dentro de cada ítem: para los ítems positivos estas distancias eran mayores, con lo que discriminaban moderadamente, mientras que para los ítems negativos las distancias eran menores, con lo que discriminaban mucho más. Estos resultados planteaban la cuestión de si este cuestionario debía ser considerado como una escala unidimensional o si debería ser dividido en dos componentes, uno negativo y otro positivo. Al mismo tiempo, planteaban la posibilidad de que el sistema de puntuación mediante números enteros que se le asigna habitualmente a las categorías de respuesta de los ítems del GHQ, no fuera el más adecuado.

La importancia que tiene el sistema de puntuación de las categorías de las escalas de respuesta viene dada también por la polémica a la hora de asignar números a dichas categorías cuando hay ítems positivos y negativos en un mismo cuestionario. Por ejemplo, el GHQ presenta, como la mayoría de cuestionarios, ítems positivos y negativos, para los cuales se han propuesto diferentes sistemas de puntuación. Goodchild y Duncan-Jones (1985) han propuesto el sistema 0, 0, 1, 1 para los ítems positivos y el sistema 0, 1, 1, 1 para los ítems negativos. Esta propuesta fue realizada porque, según estos autores, la opción de respuesta “no más de lo habitual” (que figura en segundo lugar) debe ser tratada como un indicador de enfermedad crónica más que como un indicador de buena salud. Según este

sistema de puntuación, la puntuación total obtenida por un sujeto es mejor predictor de mala salud mental. Pero también estos autores hacen otras propuestas de sistemas de puntuación para el GHQ, todo lo cual lleva a la conclusión de la importancia que tiene el sistema de puntuación de las categorías de las escalas de respuesta.

Todos estos estudios ponen de manifiesto que ha sido estudiado poco estudiado el sistema de puntuación en las escalas de respuesta ordinales, así como si en dichas escalas de respuesta las categorías se presentan realmente ordenadas, como se asume desde un principio. Si se pretende medir una conducta que no es directamente observable, para lo cual se administra un test o un cuestionario, se asume que la puntuación numérica obtenida refleja el auténtico valor que presenta el sujeto en el rasgo. Si el valor asignado a cada categoría de respuesta no es el adecuado, bien por el significado que los sujetos otorgan a dicha categoría, o bien porque el orden no es el que se ha asumido desde un principio, se vulnera el isomorfismo que ha de existir entre los sistemas empírico y numérico, haciendo que el instrumento de medida nos ofrezca un resultado confundente por su falta de validez.

CAPÍTULO 2.

MÉTODO Y OBJETIVOS

2.1. Objetivos

Como se ha visto con anterioridad, no hay muchos trabajos en los que se estudie si realmente las escalas de respuesta habitualmente utilizadas están ordenadas como se supone que deberían estar, si lo que se pretende es representar correctamente el orden existente en el sistema empírico. La ruptura del isomorfismo entre ambos sistemas relacionales (numérico y empírico), puede afectar a la fiabilidad y a la validez del instrumento haciendo fallido el proceso de inferencia en la medición.

Puesto que ha sido bastante cuestionada la puntuación asignada al punto medio en las escalas de respuesta debido a los distintos significados que se le pueden otorgar a dicho punto medio, siendo especialmente cuestionada la puntuación asignada a la alternativa “?”, el presente trabajo tiene un doble objetivo. Por una parte, se pretende comprobar si el isomorfismo entre los sistemas relacionales numérico y empírico se

mantiene para escalas de respuesta con distintos puntos medios. Es decir, se pretende comprobar si realmente las categorías de respuesta de la escala se presentan ordenadas tal y como asumen los respectivos sistemas de puntuación. Pero además se pretende estudiar el significado que los sujetos otorgan a los distintos puntos medios de dicha escala de respuesta.

El primer objetivo (estudiar si las categorías de respuesta se presentan de manera ordenada), se ha llevado a cabo poniendo a prueba el funcionamiento de tres puntos medios diferentes mediante un modelo politómico de teoría de respuesta al ítem. Por una parte, se han sometido a estudio los dos puntos medios que aparecen con mayor asiduidad en el 16 PF (Cattell et al., 1970; 1993). Estos puntos medios son “término medio” y “no estoy seguro”, y lo que se pretende es comprobar si las categorías de la escala de respuesta se presentan ordenadas tal y como se supone que deben estar si realmente se trata de una escala de respuesta tipo Likert.

También se ha tomado un tercer punto medio como objeto de estudio. Se trata de la interrogación (?). Se ha elegido también este tercer punto medio no sólo porque en la literatura hay bastante confusión acerca de su significado y, por consiguiente, acerca del sistema de puntuación más adecuado para el mismo, sino porque además es el punto medio que aparece en todos los ítems de la última versión del 16 PF (Cattell et al., 1993).

Si se comprobara que el funcionamiento de estos tres puntos medios no es el esperado, esto indicaría que el sistema de puntuación empleado, que asume un orden establecido para las categorías de respuesta de la escala, no sería el adecuado. Esto afectaría a la fiabilidad y validez de los factores

cuyas cuestiones presentaran dicho formato de respuesta y, por consiguiente, el proceso de inferencia se vería afectado. Todo ello implicaría que los resultados obtenidos a partir de un sistema de puntuación inadecuado harían llegar a conclusiones inadecuadas, lo cual es bastante grave especialmente si se tiene en cuenta el amplio uso que se hace de este cuestionario en los distintos ámbitos profesionales de la Psicología.

Pero hay un segundo objetivo, que es el de estudiar el significado que se le otorga al punto medio de las escalas respuesta. En el caso de que algunas categorías de respuesta no se presenten ordenadas tal y como se espera, estudiar dicho significado puede ayudar a comprender por qué puede ocurrir algo así, y también puede ayudar a indicar qué sistema de puntuación resultaría ser el más adecuado en ese caso.

2.1.1. La medición en personalidad

El objetivo de la medición en personalidad es casi siempre la medición de los rasgos de personalidad (Wiggins y Pincus, 1992). Aunque los motivos, intenciones, creencias, estilos y estructuras constituyen el principal foco de investigación para muchos investigadores, la investigación acerca de los rasgos de personalidad es predominante.

El Cuestionario Factorial de Personalidad 16 PF (Cattell et al., 1970) ha llegado a ser, para algunos investigadores, un sistema de medición de la personalidad más que un único instrumento de medida. Es un instrumento de medida ampliamente utilizado para medir la personalidad adulta en

individuos normales, en términos de 16 factores razonablemente independientes (Krug y Johns, 1990).

Desde los años 50, el 16 PF se ha utilizado ampliamente en la selección de personal, por lo que existe un gran número de trabajos referidos a los correlatos que las características de personalidad presentan con las preferencias ocupacionales, el rendimiento en el trabajo, el absentismo y la satisfacción laboral. Además se utiliza en ámbitos clínicos como complemento de la información obtenida por instrumentos más clínicos, como el MMPI (Hathaway y McKinley, 1967).

Más recientemente, la investigación con el 16 PF se ha centrado en las diferencias de personalidad existentes en ciertos ámbitos como son los intereses académicos y los tipos e intereses ocupacionales (Bolton, 1985; Walter, 1985). Por ejemplo, se ha podido comprobar que la extraversión y sus componentes permiten diferenciar consistentemente entre personas cuya ocupación incluye amplio contacto con otras personas frente a ocupaciones que ofrecen oportunidades limitadas de interacción.

Otras áreas de investigación en las que se ha empleado el 16 PF han sido el abuso de sustancias y su tratamiento, el estudio de las características de personalidad en las mujeres que sufren malos tratos y en los padres que abusan de sus hijos, y la investigación centrada en la terapia de pareja.

Aunque ha sido el inventario de personalidad más ampliamente utilizado en diversos campos aplicados, así como en investigación, el 16 PF ha sido objeto a menudo de duros ataques en los que se cuestiona la

fiabilidad y validez del mismo (Saville y Blinkhorn, 1981) así como su estructura factorial (Howarth y Browne, 1971), mientras que otros trabajos lo defienden incondicionalmente (Cattell, Eber y Delhees, 1968; Cattell y Krug, 1986).

Así pues, el 16 PF es un instrumento de medida que se utiliza ampliamente en diversas áreas del campo profesional, así como en la investigación. Sin embargo, los distintos estudios a los que ha sido sometido han ofrecido resultados de muy diversa índole. Además, incluye escalas de respuesta con alternativas centrales cuyo significado ha sido cuestionado y debatido, lo que podría dar lugar a que las categorías de respuesta no se presentaran ordenadas tal como asume el sistema de puntuación que emplea. Por todo ello se ha tomado como instrumento de medida objeto de estudio el Cuestionario Factorial de Personalidad 16 PF (Cattell et al., 1970; 1993) en su versión española (Seisdedos, 1989; 1995).

2.2. Muestra

Para realizar este trabajo se han empleado dos muestras, cada una para un estudio diferente. La muestra del primer estudio ha sido facilitada por TEA, y consta de 633 sujetos, los cuales presentan una media de edad de 21,62 años, con una desviación típica de 3,57. La edad mínima es de 16 años y la edad máxima es de 53 años. En la muestra hay un 50,4 % de hombres ($n = 319$) y un 49,6 % de mujeres ($n = 314$).

La muestra del segundo estudio ha sido recogida en su mayor parte entre estudiantes de Psicología, y consta de 816 sujetos de los cuales un 88,1 % (n = 721) son estudiantes, mientras que un 9 % (n=74) son trabajadores en activo. La media de edad es de 22,19 años, con una desviación típica de 6,18. La edad mínima es de 15 años y la edad máxima es de 62 años, con un 21,9 % de hombres (n = 179) y un 78,1 % de mujeres (n = 637). El 10 % de la muestra ha cursado estudios primarios, el 70,2 % ha estudiado formación profesional de segundo grado, B.U.P. o bachiller, el 13,2 % posee una titulación universitaria media y el 5,4 % son licenciados, técnicos superiores o doctores.

2.3. Variables e instrumentos de medida

El instrumento de medida que se incluye en el primer estudio es la versión cuarta del Cuestionario Factorial de Personalidad 16 PF (Cattell et al., 1970), y en el segundo estudio se incluye este cuestionario y además la versión quinta del mismo (Cattell et al., 1993), el Cuestionario de Ansiedad Estado-Rasgo (Spielberger et al., 1970), en su versión de ansiedad-rasgo y la Escala de Autoestima de Rosenberg (Rosenberg, 1965).

La versión cuarta del 16 PF (Cattell et al., 1970) consta de seis formas diferentes: A, B, C, D, E y F. Las dos primeras formas son aplicables a adolescentes y adultos con un nivel cultural equivalente al de enseñanza media, las formas C y D requieren menor nivel cultural, y las formas E y F se aplicarían a jóvenes o adultos menos inteligentes, menos competentes o

con deficiencias culturales. En este trabajo se ha estudiado la forma A, que consta de 184 ítems. Para medir cada uno de los dieciséis factores se emplean 10 ó 13 ítems, según el factor. Cada uno de los ítems presenta una escala de respuesta tipo Likert de tres puntos, y de las tres categorías de respuesta de la escala, el extremo que se presenta en dirección al constructo a medir recibe 2 puntos, el otro extremo recibe 0 puntos, y el punto medio recibe 1 punto.

Sin embargo, las categorías de respuesta no son siempre las mismas para cada ítem, ni siquiera para los ítems de un mismo factor. Por esta razón, de los dieciséis factores que mide este cuestionario se ha trabajado con los que presentaban al menos tres ítems con una escala de respuesta que contuviera el mismo punto medio. Estos factores son los siguientes:

Factor C (estabilidad emocional). De los trece ítems que lo integran, se ha trabajado con los cinco ítems que presentan una escala de respuesta cuyo punto medio es “no estoy seguro” (ítems 23, 42, 61, 98 y 174).

Factor E (dominancia). De los trece ítems que lo integran, se ha trabajado con los siete ítems que presentan una escala de respuesta cuyo punto medio es “término medio” (ítems 43, 81, 100, 138, 156, 157 y 175).

Factor F (animación). De los trece ítems que lo integran, se ha trabajado con los cinco ítems que presentan una escala de respuesta cuyo punto medio es “término medio” (ítems 44, 82, 101, 158 y 177), y con otros siete que presentan una escala de respuesta cuyo punto medio es “no estoy seguro” (ítems 7, 25, 45, 63, 64, 120 y 139).

Factor G (atención a normas). De los diez items que lo integran, se ha trabajado con los cinco items que presentan una escala de respuesta cuyo punto medio es “término medio” (items 65, 102, 140, 159 y 178).

Factor H (atreimiento). De los trece items que lo integran, se ha trabajado con los nueve items que presentan una escala de respuesta cuyo punto medio es “término medio” (items 28, 47, 66, 84, 103, 122, 123, 160 y 179).

Factor L (vigilancia). De los diez items que lo integran, se ha trabajado con los cinco items que presentan una escala de respuesta cuyo punto medio es “término medio” (items 11, 87, 106, 125 y 143).

Factor M (abstracción). De los trece items que lo integran, se ha trabajado con los cinco items que presentan una escala de respuesta cuyo punto medio es “término medio” (items 12, 88, 107, 163 y 183), y con otros tres que presentan una escala de respuesta cuyo punto medio es “no estoy seguro” (items 144, 145 y 164).

Factor N (privacidad). De los diez items que lo integran, se ha trabajado con los cuatro items que presentan una escala de respuesta cuyo punto medio es “término medio” (items 51, 70, 89 y 127), y con otros tres que presentan una escala de respuesta cuyo punto medio es “no estoy seguro” (items 13, 108 y 165).

Factor O (aprensión). De los trece items que lo integran, se ha trabajado con los siete items que presentan una escala de respuesta cuyo punto medio es “término medio” (items 14, 33, 53, 90, 128, 166 y 185).

Factor Q2 (autosuficiencia). De los diez items que lo integran, se ha trabajado con los tres items que presentan una escala de respuesta cuyo punto medio es “término medio” (items 17, 92 y 149).

Factor Q3 (perfeccionismo). De los diez items que lo integran, se ha trabajado con los tres items que presentan una escala de respuesta cuyo punto medio es “término medio” (items 56, 93 y 112), y con otros cuatro que presentan una escala de respuesta cuyo punto medio es “no estoy seguro” (items 18, 75, 131 y 150).

Factor Q4 (tensión). De los trece items que lo integran, se ha trabajado con los nueve items que presentan una escala de respuesta cuyo punto medio es “término medio” (items 19, 38, 57, 95, 113, 114, 132, 133 y 170).

La versión quinta del 16 PF (Cattell et al., 1993), presenta también 184 items, a partir de los cuales se pueden obtener puntuaciones en las mismas dieciséis escalas primarias de personalidad. Estos nuevos items se extrajeron de las ediciones anteriores en función de si presentaban una alta correlación con su misma escala y de si la relación era menor con otras escalas. De los items que aparecen en esta última versión, hay algunos items que no han sufrido cambios respecto a los de versiones anteriores, otros han

sufrido pequeños cambios, otros se han modificado bastante y otros son nuevos.

La escala de respuesta de los items es también una escala tipo Likert de tres puntos que, a diferencia de las escalas de respuesta de la versión anterior, sí presenta el mismo punto medio para todos los items, en este caso la interrogación (?), si bien los extremos de la escala pueden ser distintos, incluso para los items de un mismo factor. Además, el sistema de puntuación de las categorías de la escala es igual al de la versión cuarta (2, 1, 0).

De este cuestionario se ha trabajado con un grupo de items del factor H (atrevimiento) que apenas ha sufrido cambios respecto a la versión anterior, por lo que han sido fácilmente identificados como pertenecientes a este factor. Esta identificación ha tenido que realizarse estudiando los items del 16 PF-5 uno a uno, ya que el sistema de corrección del nuevo cuestionario se realiza mediante un programa informático que ofrece la puntuación directa del sujeto en cada factor a partir de la introducción de las respuestas dadas por el sujeto a los items del cuestionario. Este procedimiento no permite conocer los items que integran cada uno de los factores que mide el instrumento como ocurría en la versión anterior, en la cual la obtención de las puntuaciones directas de cada factor se realizaba mediante una plantilla de corrección manual.

Los otros dos instrumentos de medida empleados se utilizaron en el segundo estudio. Son instrumentos que miden sendas variables que se utilizarán como criterios para estudiar la validez de los factores del 16 PF. Estas variables son la ansiedad como rasgo, que presenta correlaciones

significativas con los factores C- (inestabilidad emocional), E- (sumisión), F- (sobriedad), N+ (privacidad), O+ (aprensión) y Q4+ (tensión) (Seisdedos, 1988), y la autoestima, que se relaciona significativamente con casi todos los factores del 16 PF (Cattell et al., 1993).

El Cuestionario de Ansiedad Estado-Rasgo (Spielberger et al., 1970), presenta dos escalas: una mide ansiedad-estado y la otra ansiedad-rasgo. Para este trabajo se ha tomado la escala de ansiedad-rasgo, compuesta por 20 items que se responden mediante una escala tipo Likert de cuatro puntos que oscila entre “1. Casi nunca” y “4. Casi siempre”, donde a mayor puntuación, mayor nivel de ansiedad-rasgo. El coeficiente alfa de Cronbach para la escala ha sido de 0,89, y la media y desviación típica para la escala han sido de 43,07 y 9,98, respectivamente.

La Escala de Autoestima de Rosenberg (Rosenberg, 1965), que ofrece una puntuación total en autoestima, presenta 10 items con una escala de respuesta tipo Likert de cuatro puntos que oscila entre “1. Estoy muy de acuerdo” y “4. Estoy muy en desacuerdo”. Cuanto mayor es la puntuación obtenida, mayor es el grado de autoestima reflejado. El coeficiente alfa de Cronbach para la escala ha sido de 0,77, y la media y desviación típica para la escala han sido de 19,07 y 4,03, respectivamente.

2.4. Análisis

En este trabajo se han realizado tres tipos de análisis: los análisis clásicos de validez, un análisis del funcionamiento de los items mediante un modelo politómico de teoría de respuesta al ítem y, un estudio de las razones dadas por los sujetos respecto al significado del punto medio en las escalas de respuesta. Los análisis de validez, así como los correspondientes al estudio de las razones, se han realizado con el paquete estadístico SPSS 6.0 para Windows.

Como se explicará con más detalle en el punto siguiente, se han llevado a cabo dos estudios. En el primero, inicialmente se ha comprobado la unidimensionalidad de cada factor del 16 PF objeto de estudio, ya que el modelo politómico escogido para estudiar las escalas de respuesta de dicho cuestionario es un modelo unidimensional. Para ello se ha realizado un análisis factorial confirmatorio de la matriz de correlaciones policóricas entre los items de cada factor mediante el programa LISREL 8 (Joreskog y Sörbom, 1993).

Tras comprobar la existencia de unidimensionalidad, se ha aplicado el modelo de categorías nominales de Bock (Bock, 1972), implementado en el programa MULTLOG (Thissen, 1988). Se ha escogido este modelo en lugar de un modelo para formato de respuesta ordinal precisamente porque se quiere comprobar si realmente las categorías de respuesta siguen el orden que se asume en un principio.

Asumiendo que las respuestas a los ítems dependen del valor del sujeto en el rasgo latente, la función de cada categoría de respuesta, es decir, la probabilidad de una respuesta en la categoría h del ítem i , se puede expresar en los siguientes términos:

$$P_{ih}(\theta) = e^{(a_{ih}\theta + b_{ih})} / \sum e^{(a_{ih}\theta + b_{ih})}$$

Un modelo politómico para formatos de respuesta ordinales no permite comprobar si realmente las categorías se presentan ordenadas, sino que lo asume. El modelo nominal de Bock, sin embargo, no asume a priori dicho orden y además permite comprobar si las categorías de respuesta están ordenadas. Se trata de un modelo logístico de dos parámetros que permite estimar un parámetro a y un parámetro b para cada categoría de respuesta (k) de cada ítem (j) (a_{jk} y b_{jk}). Samejima (1972) mostró que si los valores del parámetro a se presentan ordenados, entonces las correspondientes probabilidades acumulativas de las categorías están ordenadas también. O lo que es lo mismo, si las funciones de las categorías de respuesta están ordenadas, en el sentido de que la diferencia entre responder a una categoría superior de respuesta frente a una categoría inferior se incrementa con el valor del rasgo latente q , entonces los correspondientes valores del parámetro a están ordenados, lo cual indicaría una ordenación de las categorías de respuesta.

Esto sugiere que utilizando el modelo nominal podemos comprobar si existe un orden entre las categorías de respuesta que justificaría el análisis posterior de las mismas utilizando un modelo para categorías de respuesta

ordinales, puesto que entonces sí se podría asumir el orden empírico de dichas categorías (Bock, 1997).

A partir de los resultados obtenidos, en el segundo estudio se han realizado los mismos análisis que fueron realizados en el primer estudio pero, puesto que se incluyeron variables dicotómicas, se evaluó el ajuste de estos datos a un modelo dicotómico de la teoría de respuesta al ítem, concretamente al modelo logístico de dos parámetros, implementado en el programa BILOG (Mislevy y Bock,).

Además se ha estudiado la validez criterial de los factores del 16 PF incluidos en el estudio y se ha llevado a cabo también un estudio de las razones o significados que los sujetos daban a cada uno de los puntos medios.

2.5. Diseño

Este trabajo presenta dos partes diferenciadas. En primer lugar se ha realizado un estudio con la muestra facilitada por TEA, con el objeto de estudiar el funcionamiento de los puntos medios “término medio” y “no estoy seguro” en los ítems del 16 PF-4 mediante un modelo politómico de la teoría de respuesta al ítem, como se ha indicado anteriormente.

A partir de los resultados obtenidos se ha realizado un segundo estudio cuyo diseño se especifica a continuación, y con el que se pretende

confirmar si los resultados del primer estudio vuelven a aparecer en otra muestra distinta. Además, se pretende estudiar el funcionamiento de un tercer punto medio que aparece en el 16 PF-5, la interrogación (?), junto con el significado que los sujetos otorgan a cada uno de los puntos medios.

Este segundo estudio se ha llevado a cabo con la muestra de 816 sujetos. Para realizarlo se elaboró una batería de items que debía responder cada sujeto y que constaba de los siguientes elementos: el cuestionario de ansiedad STAI en su versión de ansiedad-rasgo, la escala de autoestima de Rosenberg, una serie de variables demográficas y quince items del 16 PF, los cuales se repetían cuatro veces, cada vez con una escala de respuesta diferente.

Por otra parte, y con el objeto de estudiar el significado que los sujetos atribuyen a cada uno de los puntos medios en las escalas de respuesta, al final de la batería se ofrecía a los sujetos una lista de posibles razones por las cuales se podía haber elegido el punto medio (cualquiera de los tres sometidos a estudio) como respuesta a alguno de los items que presentaban esta opción de respuesta, y se les pedía que indicaran la razón por la que habían respondido esta categoría, en el caso de que lo hubieran hecho. El procedimiento de elaboración de la batería se detalla a continuación.

Los quince items del 16 PF cuyo enunciado se presentaba repetido en la batería fueron los siguientes: por una parte cinco items del factor C (items 23, 42, 61, 98 y 174) y otros cinco items del factor Q4 (items 19, 57, 113, 132 y 170), incluidos en el 16 PF-4, y por otra parte cinco items del

factor H incluidos en el 16 PF-5 (items 71, 107, 135, 137 y 167), los cuales también aparecen en la versión cuarta de este cuestionario con pequeñas modificaciones.

Para comprobar si las categorías de respuesta de las escalas presentaban un funcionamiento distinto en función del punto medio de las mismas, a cada uno de los items se le cambió el punto medio que aparecía en la escala de respuesta original por cada uno de los otros puntos medios. Es decir, los items del factor C, que originalmente aparecen en el 16 PF-4 con una escala de respuesta cuyo punto medio es “no estoy seguro”, también se presentaban a cada sujeto que respondía la batería con el punto medio “término medio” y con el punto medio “?”; los items del factor Q4, que originalmente aparecen en este cuestionario con una escala de respuesta cuyo punto medio es “término medio”, también se presentaban a cada sujeto con el punto medio “no estoy seguro” y con el punto medio “?”; y los items del factor H, que originalmente aparecen en el 16 PF-5 con una escala de respuesta cuyo punto medio es “?”, se presentaban de nuevo a cada sujeto con el punto medio “no estoy seguro” y con el punto medio “término medio”. Además, para comprobar si los items con una escala de respuesta de tres puntos ofrecen más información que cuando la escala de respuesta es dicotómica, se presentó una cuarta forma de cada uno de los quince items originales cuya escala de respuesta era dicotómica, es decir, se eliminó el punto medio de dicha escala de respuesta.

De este modo se elaboraron cuatro listas con los quince items del 16 PF sometidos a estudio, pero con una escala de respuesta diferente en cada lista. Por lo tanto, cada sujeto que respondía la batería elaborada había de

responder cuatro veces a los quince items del 16 PF sometidos a estudio, aunque con escalas de respuesta diferentes.

Para designar el orden de presentación de cada una de las cuatro escalas de respuesta (dicotómica y con cada uno de los puntos medios: “?”, “término medio” y “no estoy seguro”), se procedió como sigue.

En cada una de las cuatro listas creadas, los cinco items que integraban cada factor presentaban una única escala de respuesta, la cual era diferente a la escala de respuesta de los otros dos factores. Para la primera lista la escala de respuesta de los items de cada factor se asignó al azar, pero cuidando de que fuera distinta para cada factor. Para cada una de las tres listas restantes, la escala de respuesta fue diferente para cada factor, realizando de nuevo esta asignación al azar, pero manteniendo la restricción de que fuera distinta a la escala de respuesta que aparecía en la lista anterior. En definitiva, en cada una de las cuatro listas elaboradas con los items seleccionados del 16 PF, los items de cada factor presentaban una escala de respuesta distinta a la de los items de los restantes factores, al mismo tiempo que era distinta en cada una de las cuatro listas.

Concretamente, en la primera lista se sortearon los cuatro formatos de respuesta, de manera que a los items del factor C le correspondieron como escala de respuesta aquella que contiene la “?”, a los del factor H la escala con el “término medio” y a los del factor Q4 la escala de respuesta dicotómica. De este modo se consiguió que las escalas de respuesta fueran distintas para cada factor en la primera lista. Para que también fueran distintas entre sí en las tres listas restantes, y que al mismo tiempo fueran

diferentes a las de la lista anterior, se volvió a aleatorizar para C las posibles escalas de respuesta, de entre las tres que quedaron en su caso, de modo que en la lista segunda le correspondió una escala de respuesta con punto medio igual a “no estoy seguro”, en la lista tres la escala de respuesta dicotómica y, por último, en la lista cuarta, la última escala de respuesta que quedó (punto medio = “término medio”). Con los factores H y Q4 se procedió del mismo modo, pero cuidando que dentro de cada lista las escalas de respuesta fueran distintas para cada factor.

A continuación, se aleatorizó la presentación de los quince ítems en cada una de las cuatro listas y, para minimizar todo lo posible cualquier efecto de recencia de los ítems, puesto que son los mismos aunque con escalas de respuesta diferentes, se presentó entre cada lista una serie de cuestiones, cuyo orden de presentación fue elegido de manera aleatoria. Entre la primera y la segunda se presentaron una serie de variables demográficas, entre las cuales se incluía el sexo, la edad y el nivel de estudios. Entre la segunda y la tercera se incluyó la escala de autoestima de Rosenberg, y entre la tercera y la cuarta lista se introdujo el cuestionario de ansiedad STAI en su versión de ansiedad-rasgo.

De este modo, tras una página de instrucciones acerca de la cumplimentación de la batería, a los sujetos se les presentaban todas las cuestiones en el siguiente orden y en folios separados: lista primera de ítems del 16 PF, variables demográficas, lista segunda, primer criterio (escala de autoestima), lista tercera, segundo criterio (escala de ansiedad-rasgo) y lista cuarta. Por último, se les presentó una última página con instrucciones referentes a los significados que podía tener el punto medio para los sujetos,

como se ha mencionado al principio, solicitando que indicaran la razón por la que habían elegido la opción del punto medio en el caso de que lo hubieran hecho. Para ello se les ofreció una lista en la que figuraba una serie de significados que podía tener el punto medio, y para responder a dicha cuestión se habilitó un espacio, indicado por una línea continua, junto a cada uno de los items del 16 PF. En el Anexo 1 figura una versión de la batería elaborada.

Tras elaborar esta versión de la batería, y para controlar los efectos que el orden de presentación de las escalas de respuesta pudieran tener a la hora de responder a los items del 16 PF, se contrabalanceó el orden de presentación de cada una de las listas en forma de cuadrado latino, el cual fue seleccionado al azar de entre todos los posibles. De este modo, se obtuvo un total de cuatro versiones de esta batería, de forma que en cada versión los sujetos empezaban respondiendo a los items de cada factor con una escala de respuesta diferente.

En este caso, las listas que configuraron la primera versión de la batería se denominaron a, b, c, d. Cada una de las versiones siguientes de la batería empezó por una lista distinta, elegida al azar. De este modo, la segunda versión vino dada por las listas c, a, d, b; la tercera versión por las listas b, d, a, c; y por último, la cuarta versión siguió el orden d, c, b, a.

CAPÍTULO 3.

RESULTADOS: ESTUDIO PRIMERO

Como se ha visto con anterioridad, no hay muchos trabajos en los que se estudie si realmente para las escalas de respuesta en las que se asume que las categorías de respuesta se presentan ordenadas esto es realmente así. Si lo que se pretende es representar correctamente el orden existente en el sistema empírico, la ruptura del isomorfismo entre ambos sistemas relacionales (numérico y empírico), puede afectar a la fiabilidad y a la validez del instrumento haciendo fallido el proceso de inferencia en la medición.

Puesto que ha sido bastante cuestionada la puntuación asignada al punto medio en las escalas de respuesta debido a los distintos significados que se le pueden otorgar a dicho punto medio, siendo especialmente cuestionada la puntuación asignada a la alternativa “?”, el presente trabajo pretende comprobar, por una parte, si el isomorfismo entre los sistemas relacionales numérico y empírico se mantiene para escalas de respuesta con distintos puntos medios. Es decir, se pretende comprobar si realmente las

categorías de respuesta de la escala se presentan ordenadas tal y como se asume en un principio. Pero por otra parte se pretende estudiar el significado que los sujetos otorgan a los distintos puntos medios de dicha escala de respuesta.

Este primer estudio se ha realizado con la muestra ofrecida por TEA (n = 633), y se han sometido a estudio los dos puntos medios que aparecen con mayor asiduidad en el 16 PF (Cattell et al., 1970; 1993) mediante un modelo politómico de teoría de respuesta al ítem. Estos puntos medios son “término medio” y “no estoy seguro”, y lo que se pretende es comprobar si las categorías de la escala de respuesta se presentan ordenadas tal y como se supone que deben estar si realmente se trata de una escala de respuesta tipo Likert.

Para comprobar la existencia de un orden entre las categorías de la escala de respuesta se ha llevado a cabo un análisis de cada factor mediante un modelo politómico de la teoría de respuesta al ítem, concretamente, el modelo nominal de Bock (Bock , 1972). Puesto que se trata de un modelo que asume la unidimensionalidad del cuestionario, el primer paso ha sido comprobar la unidimensionalidad de cada uno de los factores del 16 PF incluidos en el estudio mediante un análisis factorial confirmatorio llevado a cabo con el programa LISREL 8 (Joreskog y Sörbom, 1993). Los resultados se presentan a continuación.

3.1. Análisis factorial confirmatorio

Los análisis factoriales llevados a cabo para comprobar la unidimensionalidad de cada uno de los factores del 16 PF fueron realizados para cada subgrupo de items con una escala de respuesta que contuviera el mismo punto medio. Así, si dentro de algún factor había dos grupos de items con una escala de respuesta cuyo punto medio era diferente, la comprobación de la unidimensionalidad fue realizada para cada uno de estos subgrupos de items, siempre que estuviera integrado por al menos tres items. Los factores que contenían dos subgrupos fueron los siguientes: F, M, N y Q3.

En total se han llevado a cabo dieciséis análisis factoriales confirmatorios a partir de la matriz de correlaciones policóricas entre los items, y de todos ellos los resultados han sido satisfactorios únicamente para los siguientes factores: C, E, F (subgrupo con punto medio = “no estoy seguro”), G, H, L, O y Q4. En el caso de los factores F y O, los resultados fueron satisfactorios después de eliminar sendos items. Concretamente, en el factor F el modelo no alcanzaba la convergencia y se eliminó el ítem 7 por presentar una saturación factorial no significativa ($\lambda_7 = 0.16$)¹, quedando seis items. Como resultado se llegó a una solución factorial en la que los índices de ajuste resultaron satisfactorios. Respecto al factor O, se observó que el ítem 128 presentaba una saturación factorial no significativa ($\lambda_{128} = 0.01$) por lo que se procedió a su eliminación. Esto hizo que los índices de

¹ En todos los casos, el subíndice de λ hace referencia al número del ítem

ajuste del modelo para los seis ítems restantes presentaran una mejora sustancial.

En la tabla 3.1 se presentan los índices de bondad de ajuste correspondientes a los análisis factoriales de los factores del 16 PF.

Tabla 3.1. Índices de bondad de ajuste para cada uno de los factores que han presentado resultados satisfactorios

	χ^2	RMSEA *	RMSR	AGFI	NNFI
C	7.53 p=0.18	0.028 p=0.79	0.041	0.99	0.90
E	22.87 p=0.062	0.032 p=0.90	0.060	0.98	0.54
F (sin ítem 7)	22.08 p=0.0086	0.048 p=0.51	0.058	0.97	0.83
G	1.84 p=0.87	0.0 p=0.99	0.027	1.00	1.07
H	50.10 p=0.0044	0.037 p=0.91	0.056	0.98	0.94
L	6.73 p=0.24	0.023 p=0.83	0.042	0.99	0.90
O (sin ítem 128)	12.35 p=0.19	0.024 p=0.91	0.040	0.99	0.95
Q4	60.97 p=0.00020	0.045 p=0.70	0.068	0.97	0.93

* RMSEA: root mean square error of approximation (error de aproximación del promedio de los residuales cuadráticos estandarizados); RMSR: root mean square standardized residuals (raíz cuadrada del promedio de los residuales cuadráticos estandarizados); AGFI: índice de bondad de ajuste ajustado; NNFI: índice de ajuste no normado

Como puede observarse en la tabla 3.1, el índice χ^2 sólo resultó ser estadísticamente significativo para los factores F, H y Q4, si bien es un índice de bondad de ajuste que depende del tamaño de la muestra, y el NNFI únicamente presenta valores por debajo de 0.90 en los factores E (NNFI = 0.54) y F (NNFI = 0.83). Sin embargo, los índices RMSEA, RMSR y AGFI indican que todos los modelos presentan un buen ajuste ya que, para todos los factores, el valor del RMSEA es menor al punto crítico de 0.05, el del RMSR es menor a 0.07 y el del AGFI es mayor a 0.90.

Los resultados del análisis no fueron satisfactorios para los factores F (subgrupo con punto medio = “término medio”), M, N, Q2 y Q3 (ambos subgrupos). Concretamente, para el subgrupo con “término medio” como categoría intermedia del factor F, cuatro de los cinco ítems presentaron unas saturaciones factoriales que resultaron no ser significativamente distintas de cero ($\lambda_{44} = 0.07$; $\lambda_{82} = -0.04$; $\lambda_{158} = -0.05$ y $\lambda_{177} = -0.12$) y para el ítem 101, la saturación fue superior a la unidad ($\lambda_{101} = -3.46$), dando lugar a una solución mal definida.

Por otra parte, para ambos subgrupos del factor M el análisis factorial no alcanzó la convergencia en ninguno de los dos casos. Y esto mismo ocurrió en el subgrupo del factor Q3 con punto medio igual a no “no estoy seguro”.

En el factor N (subgrupo con punto medio igual a “no estoy seguro”), las correlaciones que los ítems presentaban entre sí en la matriz de correlaciones policóricas no fueron significativamente distintas de cero en dos casos ($r_{n1,n2} = 0.012$, $p > 0.05$ y $r_{n2,n3} = 0.052$, $p > 0.05$), y ninguna de las

saturaciones factoriales de los tres ítems resultó ser significativamente distinta de cero en ningún caso. En el subgrupo con punto medio igual a “término medio” (cuatro ítems), las correlaciones entre los ítems no fueron significativamente distintas de cero en ningún caso, y lo mismo ocurrió con las cuatro saturaciones factoriales.

En el factor Q2 (tres ítems) y en el factor Q3 (subgrupo con punto medio = “término medio”) las correlaciones entre los ítems tampoco fueron significativamente distintas de cero, y de las saturaciones factoriales, dos resultaron no ser significativamente distintas de cero en cada factor. Concretamente, $\lambda_{17} = 0.06$ y $\lambda_{92} = 0.20$ para el factor Q2, y $\lambda_{93} = 0.06$ y $\lambda_{112} = 0.20$ para el factor Q3. Además, en ambos factores, la tercera saturación factorial fue superior a la unidad ($\lambda_{149} = -1.27$ para Q2 y $\lambda_{56} = 1.13$ para Q3).

Debido a que el supuesto de unidimensionalidad no fue apoyado por los resultados, los factores F (subgrupo con punto medio = “término medio”), M, N, Q2 y Q3 (ambos subgrupos) fueron desestimados para la realización de posteriores análisis.

Para aquellos factores del 16 PF en los que se pudo sostener existencia de un único factor subyacente a los ítems correspondientes (factores C, E, F (subgrupo con punto medio = “no estoy seguro”), G, H, L, O y Q4), se aplicó el modelo nominal de Bock (Bock, 1972) con el objeto de comprobar si las categorías de respuesta se presentaban ordenadas tal y como asume el sistema de puntuación del 16 PF. Los resultados obtenidos se presentan en el siguiente apartado.

Para facilitar la tarea de estimación de los parámetros y la evaluación de los modelos, y considerando que el número mínimo de ítems con el mismo punto medio en un factor era de cinco, para los factores con más de cinco ítems se escogieron aquellos ítems que presentaron las saturaciones factoriales más altas. En este caso, los factores que tenían cinco ítems desde un principio eran C, G y L. Del factor E se escogieron cinco de los siete ítems (ítems 43, 81, 138, 157 y 175), de los factores F y O se escogieron cinco de los seis ítems que presentaba cada factor (ítems 25, 45, 63, 64 y 120 del factor F, e ítems 14, 33, 53, 90 y 166 del factor O) y, de los factores H y Q4 se escogieron cinco de los nueve ítems que presentaba cada factor (ítems 28, 84, 103, 123 y 160 del factor H, e ítems 19, 57, 113, 132 y 170 del factor Q4).

3.2. Análisis con el modelo nominal de Bock (Bock, 1972)

En la tabla 3.3 se presentan los índices de bondad de ajuste obtenidos tras ajustar los datos correspondientes a cada factor al modelo nominal de Bock. La evaluación del ajuste a modelos politómicos de rasgo latente no es una cuestión resuelta satisfactoriamente todavía (Reise et al., 1993). El mayor problema que tienen los índices propuestos es que son sensibles al tamaño de la muestra, de manera que casi cualquier índice rechaza la hipótesis nula de ajuste del modelo si la muestra es suficientemente grande (Hambleton et al., 1991). Uno de los índices más

utilizados es χ^2 . De hecho lo dan la mayoría de los programas utilizados en teoría de respuesta al ítem, como por ejemplo el programa MULTILOG, que ofrece un estadístico que se distribuye como χ^2 , y que es -2 veces el logaritmo de la versosimilitud estimada para el modelo. Una alternativa sencilla, fácil de calcular e interpretar y que es utilizada por diversos investigadores (Drasgow et al., 1995; Bock, 1997; Gray-Little, et al. 1997) es el índice χ^2/gl . Por esta razón, se ha tomado este índice como un indicador de la bondad de ajuste de los datos al modelo.

Como se puede observar, todos los modelos presentan un ajuste aceptable, ya que el índice χ^2/gl es menor o igual a 3 (Drasgow et al., 1995).

Tabla 3.3. Índice de bondad de ajuste χ^2/gl para cada factor, y número de ítems

	C	E	F	G	H	L	O	Q4
χ^2/gl	2.44	2.15	2.48	2.79	2.53	2.31	2.30	2.89
n ítems	5	5	5	5	5	5	5	5

A continuación se presentan los valores de los parámetros a y b estimados para cada ítem de cada factor. En la tabla 3.4 se presentan los parámetros para los factores C y F, cuya escala de respuesta presenta como punto medio la categoría de respuesta “no estoy seguro”. Como se puede observar en la tabla el orden que presentan los parámetros a para los ítems de ambos factores, no es el que se asume en un principio para las categorías

probabilidad de responder la categoría 1 son los que presentan valores medios en el rasgo.

Esto rompe el isomorfismo que debe existir entre las relaciones observadas en el plano empírico y las relaciones entre los números asignados a los sujetos a partir de sus respuestas a cada ítem.

Respecto a los factores cuyas escalas de respuesta presentan como punto medio la categoría “término medio”, se han obtenido diferentes resultados. Los parámetros *a* y *b* estimados para los ítems de dichos factores se presentan en las tabla 3.5, 3.6 y 3.7.

Tabla 3.5. Parámetros *a* y *b* estimados para los ítems de los factores H y Q4, cuyo punto medio en la escala de respuesta es “término medio”

FACTOR H				FACTOR Q4					
Categorías de respuesta				Categorías de respuesta					
1 2 3				1 2 3					
ITEM 28	a	-1.57	0.33	1.24	ITEM 19	a	-0.42	-0.12	0.54
	b	0.48	-0.61	0.13		b	0.68	-0.55	-0.13
ITEM 84	a	-1.54	0.33	1.21	ITEM 57	a	-0.44	-0.05	0.38
	b	-0.36	-0.22	0.58		b	0.52	-1.01	0.48
ITEM 103	a	-0.68	0.16	0.52	ITEM 113	a	-1.57	-0.17	1.74
	b	-0.55	-0.30	0.85		b	0.60	-0.44	-0.66
ITEM 123	a	-0.86	-0.08	0.93	ITEM 132	a	-1.03	0.40	0.63
	b	-1.06	-0.05	1.10		b	2.31	-0.97	-1.35
ITEM 160	a	-0.64	0.16	0.48	ITEM 170	a	-1.71	-0.14	1.85
	b	0.35	-1.03	0.69		b	0.25	-0.44	0.19

Como se puede comprobar en la tabla 3.5, los factores H y Q4 presentan ordenados todos los parámetros a estimados para cada uno de los items que los integran, según lo cual, los sujetos con mayor probabilidad de escoger la categoría intermedia (“término medio”) son aquellos que presentan puntuaciones intermedias estimadas en el rasgo.

Respecto a los factores L y G (ver tabla 3.6) y al respecto al factor O (ver tabla 3.7), únicamente el ítem 123 del factor L presenta los parámetros a ordenados como se asume en el sistema de puntuación del 16 PF. Los restantes items de este factor, y todos los items de los factores G y O presentan unos parámetros a cuyo orden no es tampoco el que se asumía en un principio para las categorías de respuesta, sino que se han presentado en el mismo orden que para los items de los factores C y F, es decir, el parámetro a estimado para el punto medio presenta el valor más bajo, seguido del parámetro a estimado para la categoría de respuesta 1 y, por último, del parámetro a estimado para la categoría de respuesta 3 (“2. TM”, “1. No”, “3. Sí”).

Según los resultados obtenidos en estos tres factores (L, G y O), los sujetos con menores puntuaciones estimadas en el rasgo serían los que presentarían una mayor probabilidad de escoger la categoría intermedia (“2. Término medio”) como respuesta, mientras que los sujetos que tienen más probabilidad de responder la categoría 1 serían los que presentan puntuaciones medias estimadas en el rasgo. De nuevo se rompe el isomorfismo que debe existir entre las relaciones observadas en el plano empírico y las relaciones entre los números asignados para representarlas.

Tabla 3.6. Parámetros a y b estimados para los items de los factores G y L, cuyo punto medio en la escala de respuesta es “término medio”

FACTOR G				FACTOR L					
Categorías de respuesta				Categorías de respuesta					
1 2 3				1 2 3					
ITEM 65	a	-0.9	-0.33	0.42	ITEM 11	a	0.41	-0.85	0.44
	b	0.63	-1.09	0.46		b	0.63	-0.98	0.35
ITEM 102	a	-0.01	-0.56	0.57	ITEM 87	a	0.25	-0.63	0.38
	b	-0.62	-1.34	1.96		b	0.19	-0.34	0.16
ITEM 140	a	0.23	-0.75	0.52	ITEM 106	a	0.30	-0.84	0.53
	b	-0.28	-1.07	1.35		b	1.25	-1.11	-0.14
ITEM 159	a	-0.23	-0.77	1.00	ITEM 125	a	0.87	-1.31	0.44
	b	0.15	-0.64	0.49		b	-0.11	-1.19	1.30
ITEM 178	a	-0.13	-1.22	1.35	ITEM 143	a	0.44	-1.06	0.62
	b	-0.19	-1.03	1.22		b	0.74	-1.62	0.88

Tabla 3.7. Parámetros a y b estimados para los items de los factores O y E, cuyo punto medio en la escala de respuesta es “término medio”

FACTOR O				FACTOR E					
Categorías de respuesta				Categorías de respuesta					
1 2 3				1 2 3					
ITEM 14	a	-0.21	-0.69	0.89	ITEM 43	a	-0.70	1.08	-0.38
	b	1.00	-1.43	0.43		b	0.64	-0.98	0.35
ITEM 33	a	0.12	-1.35	1.23	ITEM 81	a	-0.59	1.03	-0.44
	b	0.50	-1.29	0.79		b	0.65	-0.99	0.34
ITEM 53	a	0.04	-0.73	0.70	ITEM 138	a	-0.45	0.94	-0.49
	b	0.79	-1.05	0.27		b	1.33	-1.52	0.19
ITEM 90	a	0.10	-0.60	0.50	ITEM 157	a	-0.55	1.02	-0.47
	b	0.25	-0.95	0.70		b	1.46	-1.50	0.04
ITEM 166	a	0.11	-0.77	0.65	ITEM 175	a	-0.41	0.58	-0.17
	b	0.67	-1.03	0.36		b	-0.54	-0.13	0.67

Con respecto a los ítems del factor E (ver tabla 3.7 de nuevo) los parámetros a estimados presentan un orden diferente para distintos ítems, y en ningún caso es el que se asume en el sistema de puntuación del 16 PF. Para el ítem 138 el orden de los parámetros a correspondientes a las categorías de respuesta es 312 (“3. SÍ”, “1. NO”, “2. TM”), de modo que ninguna categoría de respuesta se presenta en el lugar que le correspondería a priori. Esto indicaría que los sujetos que presentan menores puntuaciones estimadas en el rasgo son los que tienen más probabilidad de escoger la categoría 3 (SI) como respuesta, los sujetos con mayores puntuaciones estimadas en el rasgo son los que presentan una mayor probabilidad de escoger la categoría intermedia (“2. TM”) como respuesta, y los sujetos que presentan puntuaciones medias estimadas en el rasgo son los que tienen más probabilidad de responder la categoría 1 (NO).

Para el resto de los ítems del factor E, el parámetro a estimado para la categoría de respuesta 1 (NO) aparece en primer lugar, seguido del parámetro a estimado para la categoría de respuesta 3 (SI) y, por último, el parámetro a estimado para el punto medio aparece en último lugar (“1. No”, “3. Sí”, “2. TM”). Este orden de las categorías de respuesta indica que los sujetos con más probabilidad de escoger la categoría intermedia como respuesta son en este caso aquéllos que presentan mayores puntuaciones estimadas en el rasgo, mientras que los sujetos con más probabilidad de escoger la categoría 3 de respuesta son los que presentan puntuaciones estimadas en el rasgo intermedias.

3.3. Conclusiones

En función de los resultados obtenidos, se puede decir que únicamente los factores H y Q4, cuyo punto medio en la escala de respuesta es “término medio”, son los que mantienen el isomorfismo entre las relaciones observadas en el plano empírico y las relaciones entre los números asignados a los sujetos a partir de sus respuestas, ya que son los únicos factores cuyos items presentan escalas de respuesta ordenadas tal y como asume el sistema de puntuación del 16 PF. Sin embargo, para el resto de los factores analizados (factores C, F, L, G, O y E) no se preserva este isomorfismo, ya que las escalas de respuesta de los items correspondientes no se presentan ordenadas. Por ello, las relaciones observadas en el plano empírico no están adecuadamente representadas mediante el sistema de puntuación utilizado.

Puesto que en la versión cuarta del 16 PF se ha comprobado que los items de los factores incluidos en el estudio no presentan escalas de respuesta ordenadas, se planteó la cuestión de estudiar cuál sería el funcionamiento del punto medio que figura en todos los items del 16 PF-5 (Cattell et al., 1993). En esta versión del cuestionario, las escalas de respuesta son diferentes a las que se han analizado en este primer estudio, ya que han pasado a ser, para todos los items, escalas de respuesta cuyo punto medio es “?”. Como en la versión anterior, el sistema de puntuación asume que la escala de respuesta es una escala tipo Likert, cuyas categorías se presentan ordenadas, con lo que se asume que los sujetos que responden

esta categoría de respuesta (“?”) son los que tienen valores intermedios estimados en el rasgo.

Como se indicó anteriormente, en la literatura hay bastante confusión acerca del significado de la interrogación (“?”) y, por consiguiente, acerca del sistema de puntuación más adecuado para el mismo. Por ello, se planteó la realización de un segundo estudio con el objetivo estudiar en una nueva muestra si el funcionamiento de los distintos puntos medios que aparecen en las dos versiones del 16 PF (“término medio”, “no estoy seguro” y “?”) es el esperado ya que, de no ser así, se rompería el isomorfismo que debe existir entre las relaciones observadas en el plano empírico y las relaciones del sistema numérico empleado para representarlas. Esto indicaría que el sistema de puntuación empleado, que asume un orden establecido para las categorías de respuesta de la escala, no sería el adecuado.

CAPÍTULO 4.

RESULTADOS: ESTUDIO SEGUNDO

Como se indicó al principio, el objetivo principal de este trabajo ha sido estudiar el funcionamiento de las alternativas de los items con distintas escalas de respuesta, para ver si es congruente con el sistema de puntuación empleado. A partir de los resultados obtenidos en el primer estudio, en el que se ha observado que, al menos para algunos factores, el sistema de puntuación empleado en el 16 PF-4 no parece ser el más adecuado para representar las relaciones observadas en el plano empírico, se planteó realizar un segundo estudio con diversos objetivos.

En primer lugar, se planteó realizar un segundo estudio para comprobar si los puntos medios analizados con anterioridad presentaban el mismo funcionamiento en una nueva muestra; en segundo lugar, se incluyó el estudio del punto medio que aparece en todos los items de la nueva versión del 16 PF (“?”) para comprobar si el funcionamiento del mismo es el que asume el sistema de puntuación del cuestionario; por último, se quería comprobar si los items analizados en el estudio primero presentaban un

comportamiento diferente en función de la escala de respuesta, es decir, se planteó estudiar si el funcionamiento de un ítem podía ser diferente si se presentaba con escalas de respuestas diferentes.

Para llevar a cabo estos objetivos, se elaboró una batería en la que se incluyeron algunos de los factores del estudio anterior. Se seleccionó el factor C porque es uno de los dos factores del primer estudio cuyos ítems presentan una escala de respuesta con el punto medio “no estoy seguro” (factores C y O) y porque las categorías de respuesta no se presentaron ordenadas en el estudio anterior. El factor Q4 fue elegido porque los ítems que lo integran presentan una escala de respuesta con el punto medio “término medio” y porque en el estudio previo las categorías de respuesta se presentaron ordenadas. Por último, el factor H fue incluido en este segundo estudio después de haber conseguido identificar algunos de sus ítems en el 16 PF-5, con el objeto poder estudiar el funcionamiento del nuevo punto medio incluido en todos los ítems de esta nueva versión del cuestionario (“?”).

Como también se quería comprobar si la información que ofrecían los factores cuyos ítems presentaban una escala de respuesta de tres puntos era mayor que la ofrecida por ese mismo factor pero con una escala de respuesta dicotómica, se introdujo este tipo de escala de respuesta y se calculó la función de información de cada factor en función de la escala de respuesta de sus ítems. Además, se halló la eficiencia relativa de cada factor con escala de respuesta de tres puntos respecto al mismo factor pero con escala de respuesta dicotómica.

Antes de realizar los análisis aplicando los modelos de la TRI, se contrastó la unidimensionalidad de los factores incluidos en este segundo estudio, al igual que se hizo en el estudio primero. Los resultados se presentan a continuación.

4.1. Análisis factorial confirmatorio

Como en este segundo estudio aparecen escalas de respuesta dicotómicas, el análisis factorial realizado para contrastar la unidimensionalidad se realizó con la matriz de correlaciones tetracóricas para los items con dichas escalas de respuesta, y con la matriz de correlaciones policóricas para el resto de los items.

Como se puede observar en la tabla 4.1, los índices de bondad de ajuste indican que un modelo unifactorial presenta un ajuste aceptable para todas las escalas de respuesta del factor Q4 y del factor C (“?”, “término medio”, “no estoy seguro” y escala de respuesta dicotómica).

Tabla 4.1. Índices de bondad de ajuste para cada uno de los factores, en función de la escala de respuesta de los items

	χ^2	RMSEA *	RMSR	AGFI	NNFI
C ("no estoy seguro")	29.62 p=0.000017	0.078 p=0.037	0.063	0.96	0.52
C ("??")	27.18 p=0.000053	0.074 p=0.062	0.060	0.96	0.63

(cont.)					
C (“término medio”)	23.18 p=0.00031	0.067 p=0.13	0.054	0.97	0.67
C (escala dicotómica)	20.02 p=0.0012	0.061 p=0.22	0.068	0.97	0.54
H (“no estoy seguro”)	88.42 p<0.001	0.14 p<0.001	0.010	0.91	0.83
H (“?”)	63.39 p<0.001	0.12 p<0.001	0.087	0.93	0.83
H (“término medio”)	63.03 p<0.001	0.12 p<0.001	0.088	0.93	0.83
H (escala dicotómica)	34.06 p<0.001	0.085 p=0.014	0.080	0.96	0.91
Q4 (“no estoy seguro”)	20.90 p=0.00085	0.063 p=0.19	0.053	0.98	0.98
Q4 (“?”)	14.92 p=0.011	0.049 p=0.46	0.044	0.99	0.99
Q4 (“término medio”)	11.18 p=0.048	0.039 p=0.38	0.036	0.99	0.99
Q4 (escala dicotómica)	10.83 p=0.055	0.038 p=0.70	0.047	0.99	0.99

* RMSEA: root mean square error of approximation (error de aproximación del promedio de los residuales cuadráticos estandarizados); RMSR: root mean square standardized residuals (promedio de los residuales cuadráticos estandarizados); AGFI: índice de bondad de ajuste ajustado; NNFI: índice de ajuste no normado

Sin embargo, los índices de bondad de ajuste del factor H para todas las escalas de respuesta indican que el ajuste no es satisfactorio. En los cuatro casos se encontró que el índice de modificación correspondiente a la correlación entre los errores de los ítems 135 y 137 de dicho factor era bastante alto ($IM_{\theta_{135,137}} = 67.24$ para la escala de respuesta con punto medio “no estoy seguro”; $IM_{\theta_{135,137}} = 49.79$ para la escala de respuesta con punto medio “?”; $IM_{\theta_{135,137}} = 53.93$ para la escala de respuesta con punto medio

“término medio”; y $IM_{0135,137} = 28.46$ para la escala de respuesta dicotómica)¹. Así pues, se eliminó el ítem 137, que era el que presentaba la saturación factorial más baja en cada modelo y los índices de bondad de ajuste mejoraron notablemente. Los análisis que se realizaron a partir de este momento se llevaron a cabo habiendo eliminado definitivamente este ítem del factor H (ítem 137). En la tabla 4.2 aparecen los nuevos índices de bondad de ajuste para este factor.

Tabla 4.2. Índices de bondad de ajuste para el factor H, en función de la escala de respuesta de los ítems, después de eliminar el ítem con la saturación factorial más baja

	χ^2	RMSEA	RMSR	AGFI	NNFI
H-4 ítems ("no estoy seguro")	14.13 p=0.00085	0.086 p=0.059	0.034	0.97	0.94
H-4 ítems ("?")	8.29 p=0.016	0.062 p=0.26	0.025	0.98	0.96
H-4 ítems ("término medio")	7.2 p=0.027	0.057 p=0.33	0.026	0.98	0.97
H-4 ítems (escala dicotómica)	0.40 p=0.12	0.038 p=0.058	0.021	0.99	0.99

* RMSEA: root mean square error of approximation (error de aproximación del promedio de los residuales cuadráticos estandarizados); RMSR: root mean square standardized residuals (promedio de los residuales cuadráticos estandarizados); AGFI: índice de bondad de ajuste ajustado; NNFI: índice de ajuste no normado

¹ $IM_{0x,y}$ = índice de modificación para la correlación entre los errores de los ítems x e y

4. 2. Análisis con los modelos de la teoría de respuesta al ítem

Puesto que en este estudio se incluyeron escalas de respuesta dicotómicas, se estudió el ajuste de los datos obtenidos con esta escala de respuesta a un modelo dicotómico de respuesta al ítem. En este caso se aplicó el modelo logístico de dos parámetros (Birnbbaum, 1968).

En la tabla 4.4 aparecen los índices de bondad de ajuste obtenidos tras ajustar los datos obtenidos para los ítems de cada factor con escala de respuesta politómica al modelo nominal de Bock. Como se puede observar, todos los modelos presentan un ajuste aceptable, ya que el índice $\chi^2/g.l$ es menor o igual a 3.

Tabla 4.4. Índice de bondad de ajuste $\chi^2/g.l$ para cada factor con escala de respuesta de tres puntos, y número de ítems de cada factor

	CNS	CQM	CTM	HNS	HQM	HTM	QNS	QQM	QTM
$\chi^2/g.l$	2.03	2.15	2.36	2.23	2.68	3.00	2.14	3.00	2.57
n ítems	5	5	5	4	4	4	5	5	5

CNS: factor C con punto medio “no estoy seguro”; CQM: factor C con punto medio “?”; CTM: factor C con punto medio “término medio”; HNS: factor H con punto medio “no estoy seguro”; HQM: factor H con punto medio “?”; HTM: factor H con punto medio “término medio”; QNS: factor Q4 con punto medio “no estoy seguro”; QQM: factor Q4 con punto medio “?”; QTM: factor Q4 con punto medio “término medio”

Por su parte, los parámetros a y b estimados para cada ítem dicotómico² tras la aplicación del modelo logístico de dos parámetros, y el índice de bondad de ajuste que ofrece BILOG para cada ítem se muestran en la tabla 4.5. Como puede comprobarse, todos los ítems de cada uno de los tres factores presentaron un buen ajuste al modelo, ya que el índice de ajuste que ofrece para los residuales (root mean square standardized posterior residual) presenta, en todos los casos, valores menores a 2.

Tabla 4.5. Parámetros a y b estimados, e índices de bondad de ajuste para los ítems de cada factor con escala de respuesta dicotómica

FACTOR C			
ITEM	SLOPE S.E.	THRESHOLD S.E.	RT MEAN SQUARE STD POSTERIOR RESIDUAL
C1	.609 .145*	2.085 .453*	.346
C2	1.043 .231*	-1.216 .209*	.167
C3	.553 .126*	1.287 .297*	.523
C4	.781 .172*	1.629 .311*	.304
C5	.988 .211*	-1.459 .246*	.312
FACTOR H			
ITEM	SLOPE S.E.	THRESHOLD S.E.	RT MEAN SQUARE STD POSTERIOR RESIDUAL
H1	2.161 .264*	.021 .056*	.307
H2	2.635 .416*	-.343 .054*	.560
H3	1.384 .154*	.134 .070*	.641
H4	1.484 .166*	-.072 .067*	.451

² Los parámetros a y b estimados a partir del modelo logístico de dos parámetros tienen un significado distinto al del modelo nominal de Bock.

(cont.) FACTOR Q4

ITEM	SLOPE S.E.	THRESHOLD S.E.	RT MEAN SQUARE POSTERIOR RESIDUAL	STD
Q1	.596 .093*	.411 .139*	.881	
Q2	.401 .078*	.199 .186*	.733	
Q3	4.096 1.599*	.526 .057*	.643	
Q4	1.208 .172*	-1.331 .150*	.367	
Q5	3.833 1.175*	.656 .072*	.560	

* ERROR ESTÁNDAR

A continuación se presentan los valores de los parámetros a y b estimados para cada ítem de cada factor con escala de respuesta politómica. En las tablas 4.6, 4.7 y 4.8 aparecen los valores de los parámetros estimados para los factores H, Q4 y C, en función de la escala de respuesta de los ítems. Como se puede comprobar en las tablas 4.6 y 4.7, en el factor H y en el factor Q4 los parámetros a de las categorías de respuesta se presentan ordenados en todos los casos.

Tabla 4.6. Parámetros a y b estimados para los ítems del factor H con distintas escalas de respuesta

FACTOR H									
2. Término medio					2. No estoy seguro				
Categorías de respuesta					Categorías de respuesta				
		1	2	3		1	2	3	
ITEM 71	a	-1.34	-0.04	1.38	ITEM 71	a	-1.49	0.03	1.46
	b	-0.13	0.22	-0.10		b	-0.12	-0.01	0.13
ITEM 167	a	-2.21	-0.09	2.31	ITEM 167	a	-1.90	0.09	1.81
	b	0.73	0.39	-1.12		b	0.65	0.06	-0.71
ITEM 107	a	-0.98	0.12	0.86	ITEM 107	a	-0.88	0.16	0.72
	b	-0.16	0.35	-0.18		b	-0.03	0.02	0.01

(cont.)									
ITEM	a	-1.31	0.21	1.10	ITEM	a	-1.34	0.18	1.15
135	b	-0.52	0.52	-0.10	135	b	-0.38	0.27	0.11
2. ?									
Categorías de respuesta									
		1	2	3					
ITEM 71	a	-1.27	-0.04	1.31					
	b	0.01	-0.11	0.10					
ITEM	a	-1.67	0.03	1.64					
167	b	0.63	-0.06	-0.57					
ITEM	a	-0.88	0.05	0.83					
107	b	-0.02	0	0.02					
ITEM	a	-1.19	0.09	1.10					
135	b	-0.27	0.27	0					

Tabla 4.7. Parámetros a y b estimados para los items del factor Q4 con distintas escalas de respuesta

FACTOR Q4									
2. Término medio					2. No estoy seguro				
		Categorías de respuesta					Categorías de respuesta		
		1	2	3			1	2	3
ITEM	a	-0.26	-0.13	0.39	ITEM 19	a	-0.34	-0.03	0.36
19	b	-0.06	-0.08	0.13		b	0.09	-0.33	0.23
ITEM	a	-0.17	-0.09	0.26	ITEM 57	a	-0.15	-0.07	0.22
57	b	0.10	-0.12	0.01		b	0.20	-0.30	0.09
ITEM	a	-6.70	1.43	5.27	ITEM	a	-18.55	6.55	12.00
113	b	-3.63	1.37	2.26	113	b	-9.38	3.60	5.77
ITEM	a	-0.82	-0.23	1.05	ITEM	a	-0.51	-0.08	0.58
132	b	1.15	-0.14	-1.01	132	b	0.97	-0.40	-0.57
ITEM	a	-13.38	4.93	8.46	ITEM	a	-14.70	3.72	10.98
170	b	-7.54	3.13	4.41	170	b	-7.87	2.24	5.62

(cont.)

		2. ?		
		Categorías de respuesta		
		1	2	3
ITEM 19	a	-0.22	-0.11	0.33
	b	0.12	-0.50	0.37
ITEM 57	a	-0.19	-0.02	0.22
	b	0.18	-0.30	0.37
ITEM 113	a	-8.46	-0.34	8.79
	b	-4.35	0.31	4.04
ITEM 132	a	-0.61	-0.16	0.77
	b	1.19	-0.48	-0.71
ITEM 170	a	-7.50	-0.18	7.68
	b	-4.11	0.06	4.05

Sin embargo, en el factor C los parámetros a presentan un orden distinto según la escala de respuesta (ver tabla 4.8). Concretamente, cuando la escala de respuesta tiene como punto medio la interrogación (?) y cuando el punto medio es “2. No estoy seguro”, los parámetros a estimados para los items del factor C se presentan en el orden siguiente: “2. No estoy seguro”, “1. No”, “3. Sí” (orden 213). Según estos resultados, los sujetos que tienen más probabilidad de responder el punto medio de la escala de respuesta (“?” o “2. No estoy seguro”) son aquéllos que presentan un valor más bajo en el rasgo, mientras que los sujetos que tienen más probabilidad de responder la categoría de respuesta 1 son los que presentan valores intermedios en el rasgo.

Así pues, el sistema de puntuación empleado para representar las relaciones observadas en el plano empírico no es el más adecuado para puntuar los items del factor C ya que no refleja las relaciones observadas en

el plano empírico, con lo que se rompe el isomorfismo entre ambos sistemas relacionales.

Tabla 4.8. Parámetros a y b estimados para los items del factor C con distintas escalas de respuesta

FACTOR C									
2. Término medio					2. No estoy seguro				
Categorías de respuesta					Categorías de respuesta				
1 2 3					1 2 3				
ITEM 43	a	-0.79	0.77	0.01	ITEM 43	a	0.25	-0.99	0.74
	b	-0.72	-0.38	1.10		b	-0.09	-1.08	1.17
ITEM 81	a	-0.67	0.62	-0.20	ITEM 81	a	-0.28	-0.47	0.75
	b	0.88	-0.70	-0.18		b	1.06	-0.76	-0.30
ITEM 138	a	-0.67	0.71	-0.04	ITEM 138	a	0.12	-0.51	0.38
	b	-0.86	0.35	0.51		b	-0.56	0.05	0.52
ITEM 157	a	-0.80	0.67	0.13	ITEM 157	a	0.11	-0.53	0.42
	b	-0.93	0.02	0.91		b	-0.62	-0.13	0.75
ITEM 175	a	-0.31	0.62	-0.31	ITEM 175	a	-0.25	-0.36	0.61
	b	0.87	-0.13	-0.74		b	0.95	-0.25	-0.71
2. ?									
Categorías de respuesta									
1 2 3									
ITEM 43	a	0.29	-0.82	0.52	ITEM 43	a	0.29	-0.82	0.52
	b	-0.20	-0.94	-0.31		b	-0.20	-0.94	-0.31
ITEM 81	a	-0.27	-0.55	0.82	ITEM 81	a	-0.27	-0.55	0.82
	b	1.12	-0.81	-0.31		b	1.12	-0.81	-0.31
ITEM 138	a	0.06	-0.35	0.30	ITEM 138	a	0.06	-0.35	0.30
	b	-0.48	0.01	0.48		b	-0.48	0.01	0.48
ITEM 157	a	0.08	-0.63	0.55	ITEM 157	a	0.08	-0.63	0.55
	b	-0.49	-0.33	0.82		b	-0.49	-0.33	0.82
ITEM 175	a	-0.19	-0.63	0.82	ITEM 175	a	-0.19	-0.63	0.82
	b	1.10	-0.23	-0.87		b	1.10	-0.23	-0.87

Por lo que respecta a la escala de respuesta cuyo punto medio es “2. Término medio”, los parámetros a tampoco se presentan ordenados como se asume en el 16 PF, si bien el orden con el que aparecen no es el anterior. En este caso, en el ítem 175 los parámetros a estimados para las categorías de respuesta 1 y 3 presentan el mismo valor ($a_1 = a_3 = -0.31$), frente al valor positivo del parámetro a estimado para la categoría 2 ($a_2 = 0.62$). Según este resultado, los sujetos con bajos valores en el rasgo tendrían la misma probabilidad de escoger las categorías de respuesta “1. No” y “3. Sí”. En el resto de los ítems, el orden en que se presentan los parámetros a estimados es el siguiente: “1. No”, “3. Sí”, “2. Término medio” (orden 132), según lo cual los sujetos que tienen más probabilidad de responder eligiendo la categoría de respuesta intermedia (“2. Término medio”) son aquéllos que presentan valores más altos en el rasgo, mientras que los sujetos que presentan mayor probabilidad de responder la categoría de respuesta “3. Sí” son los que presentan valores intermedios en el rasgo.

En resumen, para el factor C, cuando la escala de respuesta tiene como categoría intermedia “?” o “no estoy seguro”, los sujetos que tienen más probabilidad de escoger esta categoría como respuesta son los que tienen valores más bajos en el rasgo, mientras que cuando se trata de la escala de respuesta con “término medio” como categoría intermedia, los sujetos que tienen más probabilidad de escoger dicha categoría como respuesta son aquéllos que presentan valores más altos en el rasgo.

4.3. Conclusiones

Si se comparan estos resultados con los obtenidos para los factores incluidos en el primer estudio (ver tabla 4.8), se puede observar que los items de los factores H (atreimiento) y Q4 (tensión) presentan ordenados los parámetros a estimados tanto para la escala de respuesta del primer estudio (con punto medio = “término medio”) como para las tres escalas de respuesta del segundo estudio. Respecto al factor C (estabilidad emocional), los parámetros a estimados para la escala de respuesta del primer estudio (con punto medio = “no estoy seguro”), se presentan en el mismo orden (2. NES, 1. NO, 3. SÍ) en el segundo estudio cuando la escala de respuesta tiene como punto medio “no estoy seguro”, pero además el orden es el mismo cuando el punto medio, en lugar de “no estoy seguro” es “?”. Sin embargo, la escala de respuesta cuyo punto medio es “término medio” presenta un orden distinto (1. NO, 3. SÍ, 2. TM).

Tabla 4.8. Orden de presentación de las categorías de respuesta en cada estudio.

Factor	Primer estudio		Segundo estudio	
	Punto medio*	Orden de presentación de las categorías	Puntos medios	Orden de presentación de las categorías
H	TM	123 (NO, TM, SI)	?	123 (NO, ?, SI)
			NES	123 (NO, NES, SI)
			TM	123 (NO, TM, SI)
Q4	TM	123 (NO, TM, SI)	?	123 (NO, ?, SI)
			NES	123 (NO, NES, SI)
			TM	123 (NO, TM, SI)

(cont.)

C	NES	213 (NES,NO,SI)	?	213 (?,NO,SI)
			NES	213 (NES,NO,SI)
			TM	132 (NO,SI,TM)

* TM: término medio; NES: no estoy seguro

Estos resultados muestran que las categorías de la escala de respuesta original para los items de los factores C, H y Q4 (con punto medio NES, TM y TM, respectivamente) se presentan en el mismo orden en ambos estudios, es decir, el comportamiento de las escalas de respuesta originales es el mismo en muestras diferentes, como cabría esperar a partir de la invarianza de los parámetros. Además se observa, que el orden se preserva también para los factores H y Q4 cuando cambia la escala de respuesta. Sin embargo, el factor C no presenta las categorías de respuesta ordenadas en ninguna de las tres escalas de respuesta empleadas en este segundo estudio.

Si para los factores H y Q4 las tres escalas de respuesta se presentan ordenadas, y para el factor C no ocurre así, tal vez esto sea debido a que los sujetos consideren que el significado de cada uno de los puntos medios de las escalas de respuesta cuando responden a los items del factor C, es distinto al significado que otorgan a estos puntos medios cuando se responde a otros factores. Si esto es así, el significado que debe tener el punto medio para los sujetos, cuando responden los items de los factores H y Q4, ha de ser mayoritariamente “estar en el punto medio de un continuo”, mientras que para los items del factor C el significado otorgado a la categoría intermedia debe ser otro.

CAPÍTULO 5.

RESULTADOS: ESTUDIO DE LAS RAZONES

Los resultados descritos en el capítulo anterior han mostrado que el comportamiento de las escalas de respuesta originales de los factores C, H y Q4 (con punto medio NES, TM y TM, respectivamente) es el mismo en muestras diferentes, además de que el orden se preserva también para los factores H y Q4 cuando cambia la escala de respuesta. Sin embargo, el factor C no presenta las categorías de respuesta ordenadas en ninguna de las tres escalas de respuesta. Según esto, habría que plantearse si se puede seguir asumiendo, como se ha venido haciendo hasta ahora, que las distintas alternativas centrales de respuesta empleadas en el 16 PF tienen el significado que se les ha atribuido, es decir, término medio.

En la batería elaborada se incluyó una lista de posibles razones que podían llevar a los sujetos a responder a cada ítem con el punto medio de la escala, en el caso de que lo hubieran hecho, con el objeto de estudiar el significado otorgado al punto medio de las escalas respuesta. Quizás el estudio de las razones que los sujetos aportan cuando responden escogiendo

alguno de los puntos medios ayude a entender el funcionamiento de las escalas de respuesta y a clarificar si realmente se puede seguir asumiendo que dichos puntos medios se encuentran en mitad de un continuo.

Puesto que las categorías de respuesta para algunos ítems no se presentan ordenadas tal y como se espera, estudiar el significado que los sujetos dan al punto medio podría ayudar a comprender por qué ha podido ocurrir algo así, y también podría ayudar a indicar la adecuación del sistema de puntuación empleado.

La lista de razones que se presentó en la batería fue la siguiente:

1. No entiendo bien la cuestión que se plantea
2. No quiero revelar cuestiones personales
3. Me falta información para elegir alguna de las otras dos opciones
4. Me siento en el punto medio que hay entre ambas opciones
5. Podría contestar cualquiera de las otras dos opciones porque ambas son aplicables a mí
6. Ninguna de las otras dos opciones me describe adecuadamente
7. No me interesa la cuestión, me siento indiferente
8. Otra razón (especifica cuál en la línea correspondiente)

Esta lista se elaboró a partir de la revisión de la literatura, en la que se encontraron distintos significados que podían otorgarse al punto medio en una escala de respuesta. Cuando se trabaja con escalas tipo Likert, se asignan valores ordenados a las categorías de respuesta, asumiendo que las alternativas están espaciadas por intervalos iguales y que el sujeto que

responde atribuye a cada categoría de respuesta el mismo significado atribuido por el investigador. Sin embargo, algunas investigaciones sugieren que asumir este tipo de atribuciones puede no ser válido (Wang y Stanley, 1970). Es necesario considerar no sólo la puntuación obtenida en cada ítem, sino también considerar cómo interpreta el sujeto cada ítem y sus alternativas de respuesta (Dubois y Burns, 1975).

De acuerdo con la escasa investigación llevada a cabo (Cronbach, 1946; Edwards, 1946; Worthy, 1969; Goldberg, 1971; Kaplan, 1972) el significado esencial de la categoría “?” sería el de ambivalencia o el de indiferencia. El sujeto ambivalente es el que presenta sentimientos positivos y negativos acerca de la misma cuestión, con lo que la ambivalencia sería el resultado de un alto nivel de implicación con el objeto, mientras que el sujeto indiferente elige la interrogación porque no se siente interesado por la cuestión. En este caso, la ambivalencia estaría reflejada en la razón 5 (podría contestar cualquiera de las otras dos opciones porque ambas son aplicables a mí), y la indiferencia en la razón 7.

Pero también podría haber otras razones para escoger la “?” como respuesta (Dubois y Burns, 1975). Por ejemplo, algunas personas pueden elegir esta categoría de respuesta porque no se sienten suficientemente competentes o suficientemente informados como para elegir alguna de las demás categorías de respuesta. La competencia de los sujetos vendría dada por la razón 3 (me falta información para elegir alguna de las otras dos opciones). Otros sujetos pueden elegir la categoría “?” porque no entienden la cuestión que se les plantea (Cruickshank, 1978), y otros pueden elegirla para indicar su rechazo a revelar cuestiones personales. En este caso, estas

dos posibilidades estarían reflejadas en las razones 1 (no entiendo bien la cuestión que se plantea) y 2 (no quiero revelar cuestiones personales), respectivamente.

Por último, la razón 4 (me siento en el punto medio) indica que el sujeto considera que está situado en mitad de un continuo, que es lo que normalmente se asume cuando se incluye un punto medio en una escala de respuesta. Esta razón es la que deberían aportar en mayor medida los sujetos que respondan eligiendo el punto medio en las escalas de respuesta cuyas categorías se presentan ordenadas, mientras que para aquellas escalas de respuesta que no se presentan ordenadas las razones deberían ser básicamente diferentes, de modo que la razón 4 (me siento en el punto medio) se aportara en menor medida.

Si bien todas estas razones provinieron de los estudios realizados para estudiar el funcionamiento y significado de la categoría “?”, que ha sido uno de los puntos medios más utilizados, también es razonable plantearse si otros puntos medios diferentes, en este caso “no estoy seguro” y “término medio”, pueden presentar alguno de estos significados para los sujetos, en lugar de indicar una posición intermedia tal y como se asume habitualmente.

Por otra parte, se dejó una última opción abierta para permitir a los sujetos aportar otras razones diferentes a las de la lista. A partir del análisis de estas aportaciones se definieron tres nuevas categorías. Las tres nuevas razones que se añadieron a la lista, a partir de las aportaciones ofrecidas por los sujetos fueron las siguientes:

8. A veces me ocurre lo que se dice en el ítem
9. Depende de aspectos contextuales
10. No sé, no estoy seguro

La primera de estas tres nuevas razones indicaría demasiada poca ocurrencia de lo que se describe en el ítem como para elegir la opción 3 (SI) claramente, aunque la ocurrencia es suficiente como para que los sujetos no consideren adecuado decir que no les ocurre nunca lo que dice el ítem y elijan la opción 1 (NO), por lo que responden eligiendo la categoría la intermedia (2).

La razón 9 indica la importancia que tiene la situación o el contexto para la persona que responde, de manera que si esta situación no la perciben como claramente definida, pero sí les ocurre lo que dice el ítem, eligen la opción intermedia como respuesta porque responder 1 (NO) sería negar por completo que ocurre en ocasiones, mientras que responder 3 (SI) podrían considerarlo excesivo porque desconocen por completo la situación y sería como indicar que ocurre siempre. Por ejemplo, el ítem 167 del factor H (“Me cuesta bastante hablar delante de un grupo numeroso de personas”) no especifica qué clase de grupo es, si son conocidos o desconocidos para el sujeto, por lo que mucha gente indica que “depende del grupo, de si son conocidos o no”.

Por último, la razón 10 (no sé, no estoy seguro) incluye a aquellos sujetos que indican claramente que no saben por qué responden la categoría intermedia.

A continuación se presenta un listado global de las razones aportadas por los sujetos para cada factor (tablas 5.1, 5.2 y 5.3). Como se observa en la tabla 5.1, correspondiente al factor C, la razón 4 se aporta un mayor número de veces cuando se responde eligiendo el punto medio TM (53,17%), seguida por la escala de respuesta con punto medio “NES” (51,93%) y por la escala de respuesta con punto medio “?” (46,44%).

Tabla 5.1. Número de razones dadas para cada formato de respuesta de los items del factor C, porcentajes sobre el total de veces que se ha escogido cada punto medio de la escala de respuesta (porcentaje inferior) y porcentajes sobre el total de veces que se ha indicado cada razón (porcentaje de la derecha)

	NES	?	TM	TOTAL
1. No entiendo bien la cuestión que se plantea	23 (37,7) (2,9)	21 (34,43) (2,77)	17 (27,87) (1,89)	(45,56) (2,51)
2. No quiero revelar cuestiones personales	27 (36) (3,48)	20 (26,67) (2,64)	28 (37,33) (3,11)	(32,23) (3,08)
3. Me falta información para elegir alguna de las otras dos opciones	121 (35,17) (15,59)	118 (34,3) (15,57)	105 (30,52) (11,68)	204,17 (14,44)
4. Me siento en el punto medio que hay entre ambas opciones	403 (32,95) (51,93)	352 (28,78) (46,44)	478 (39,08) (53,17)	984,65 (50,27)

(cont.)

5. Podría contestar cualquiera de las otras dos opciones porque ambas son aplicables a mí	64 (27,12) (8,25)	96 (40,68) (12,66)	76 (32,2) (8,45)	111,64 (9,7)
6. Ninguna de las otras dos opciones me describe adecuadamente	88 (30,14) (11,34)	93 (31,85) (12,27)	111 (38,01) (12,35)	162,04 (12)

(cont.)

7. No me interesa la cuestión, me siento indiferente	16 (19,51) (2,06)	29 (35,37) (3,83)	37 (45,12) (4,12)	(21,01) (3,37)
--	----------------------	----------------------	----------------------	-------------------

8. A veces me ocurre lo que se dice en el ítem	14 (30,43) (1,8)	11 (23,91) (1,45)	21 (45,65) (2,34)	(51,58) (1,89)
9. Depende de aspectos contextuales	9 (25,71) (1,16)	10 (28,57) (1,32)	16 (45,71) (1,78)	(60,25) (1,44)
10. No sé, no estoy seguro	6 (37,5) (0,77)	5 (31,25) (0,66)	5 (31,25) (0,56)	(75,99) (0,66)
11. Otras	5 (38,46) (0,64)	3 (23,08) (0,40)	5 (38,46) (0,56)	(77,6) (0,53)
TOTAL	325,39 (31,89)	319,1 (31,15)	387,79 (36,95)	2433

NES: con punto medio “no estoy seguro”; ?: con punto medio “?”; TM: con punto medio “término medio”

Respecto al factor H (ver tabla 5.2), la razón 4 se aporta un mayor número de veces cuando se responde eligiendo el punto medio TM (57,42%), seguida por la escala de respuesta con punto medio “?” (55,24%) y por la escala de respuesta con punto medio “NES” (51,48%).

Tabla 5.2. Número de razones dadas para cada formato de respuesta de los ítems del factor H, porcentajes sobre el total de veces que se ha escogido cada punto medio de la escala de respuesta (porcentaje inferior) y porcentajes sobre el total de veces que se ha indicado cada razón (porcentaje de la derecha)

	NES	?	TM	TOTAL
1. No entiendo bien la cuestión que se plantea	8 (33,33) (1,08)	7 (29,17) (0,99)	9 (37,5) (0,93)	(78) (0,99)
2. No quiero revelar cuestiones personales	14 (29,79) (1,88)	16 (34,04) (2,27)	17 (36,17) (1,77)	(56,92) (1,95)
3. Me falta información para elegir alguna de las otras dos opciones	88 (36,36) (11,83)	66 (27,27) (9,35)	88 (36,36) (9,14)	114,69 (10,03)
4. Me siento en el punto medio que hay entre ambas opciones	383 (28,89) (51,48)	390 (29,41) (55,24)	553 (41,7) (57,42)	1065,86 (54,95)
5. Podría contestar cualquiera de las otras dos opciones porque ambas son aplicables a mí	92 (31,08) (1,09)	88 (29,73) (12,46)	116 (39,19) (12,05)	175,4 (12,27)
6. Ninguna de las otras dos opciones me describe adecuadamente	102 (32,48) (13,71)	91 (28,98) (12,89)	121 (38,53) (12,56)	180,85 (13,01)
7. No me interesa la cuestión, me siento indiferente	6 (28,57) (0,81)	7 (33,33) (0,99)	8 (38,1) (0,83)	(74,63) (0,87)
8. A veces me ocurre lo que se dice en el ítem	12 (35,29) (1,61)	10 (29,41) (1,41)	12 (35,29) (1,25)	(62,26) (1,41)
9. Depende de aspectos contextuales	28 (35,9) (3,76)	22 (28,2) (3,12)	28 (35,9) (2,91)	(22,79) (29,51)
10. No sé, no estoy seguro	1 (33,33) (0,14)	1 (33,33) (0,14)	1 (33,33) (0,1)	(87,37) (0,12)
11. Otras	10 (35,71) (1,34)	8 (28,57) (1,13)	10 (35,71) (1,04)	(64,5) (1,16)
TOTAL	294,54 (30,83)	274,57 (29,26)	455,22 (39,91)	924,33

NES: con punto medio “no estoy seguro”; ?: con punto medio “?”; TM: con punto medio “término medio”

Por último, respecto al factor Q4 (ver tabla 5.3), la razón 4 se aporta un mayor número de veces cuando se responde eligiendo el punto medio TM (54,41%), seguida por la escala de respuesta con punto medio “?” (52,52%) y por la escala de respuesta con punto medio “NES” (46,79%).

Tabla 5.3. Número de razones dadas para cada formato de respuesta de los items del factor Q4, porcentajes sobre el total de veces que se ha escogido cada punto medio de la escala de respuesta (porcentaje

inferior) y porcentajes sobre el total de veces que se ha indicado cada razón (porcentaje de la derecha)

	NES	?	TM	TOTAL
1. No entiendo bien la cuestión que se plantea	17 (48,57) (2,33)	8 (22,86) (1,39)	10 (28,57) (1,18)	(68,9) (1,62)
2. No quiero revelar cuestiones personales	13 (25,49) (1,78)	8 (12,5) (1,39)	30 (58,82) (3,33)	(50,31) (2,37)
3. Me falta información para elegir alguna de las otras dos opciones	87 (33,85) (11,9)	83 (32,2) (14,43)	87 (33,85) (10,25)	123,52 (11,93)
4. Me siento en el punto medio que hay entre ambas opciones	342 (30,92) (46,79)	302 (27,3) (52,52)	462 (41,77) (54,41)	856,29 (51,32)
5. Podría contestar cualquiera de las otras dos opciones porque ambas son aplicables a mí	98 (36,16) (13,41)	67 (24,72) (11,65)	106 (39,11) (12,48)	138,47 (12,58)
6. Ninguna de las otras dos opciones me describe adecuadamente	65 (29,41) (8,89)	65 (29,41) (11,3)	91 (41,18) (10,72)	96,09 (10,26)
7. No me interesa la cuestión, me siento indiferente	76 (70,08) (10,4)	12 (11,54) (2,09)	16 (15,38) (1,88)	(0,37) (4,83)
8. A veces me ocurre lo que se dice en el ítem	10 (25,64) (1,37)	12 (30,77) (2,09)	17 (43,59) (2)	(58,46) (1,81)
9. Depende de aspectos contextuales	18 (29,51) (2,46)	16 (26,23) (2,78)	27 (44,26) (3,18)	(38,42) (2,83)
10. No sé, no estoy seguro	5 (62,5) (0,07)	1 (12,5) (0,02)	2 (25) (0,24)	(82,33) (0,37)
11. Otras	0 (0) (0)	1 (50) (0,02)	1 (50) (0,12)	(87,14) (0,09)
TOTAL	239,47 (33,92)	195,29 (26,68)	327,68 (39,40)	728,43

NES: con punto medio “no estoy seguro”; ?: con punto medio “?”; TM: con punto medio “término medio”

En resumen, se puede ver en la tabla 5.4 que la razón 4 (me siento en el punto medio), se aporta mayor número de veces cuando se responde TM, y esto ocurre con los tres factores del 16 PF (C, H y Q4). Por lo que se refiere a la categoría de respuesta NES, la razón 4 se aporta un mayor número de veces en los factores C y H, mientras que para la “?” la razón 4 se aporta en más ocasiones en los factores H y Q4.

Tabla 5.4. Número de razones dadas para cada formato de respuesta en cada factor, y porcentaje de veces sobre el total que se ha indicado la razón 4 (ME SIENTO EN EL PUNTO MEDIO)

	FACTOR C	FACTOR H	FACTOR Q4
Término medio (TM)	53,17%	57,42%	54,41%
No estoy seguro (NES)	51,93%	51,48%	46,79%
?	46,44%	55,24%	52,52%

Estos resultados podrían ayudar a comprender el orden de los parámetros a que han sido estimados para cada escala de respuesta. Como se ha visto en el capítulo anterior, el comportamiento de las escalas de respuesta originales de los factores C, H y Q4 (con punto medio NES, TM y TM, respectivamente) es el mismo en muestras diferentes, y además este orden se preserva también para los factores H y Q4 cuando cambia la escala de respuesta. Sin embargo, el factor C no presenta las categorías de respuesta ordenadas en ninguna de las tres escalas de respuesta.

Quizá si se estudia el contenido del factor C, que mide estabilidad emocional, se pueda entender entonces por qué en este factor las escalas de respuesta no aparecen ordenadas. Recordemos que, para los items del factor C, cuando el punto medio en la escala de respuesta era “término medio” los sujetos que tenían más probabilidad de escoger esta opción como respuesta eran los que presentaban valores mayores estimados en el rasgo (orden 132; NO,SI,NES), mientras que cuando el punto medio en la escala de respuesta

era “no estoy seguro” o “?”, los sujetos que tenían más probabilidad de escoger esta opción como respuesta eran los que presentaban valores menores en el rasgo (orden 213; NES,NO,SI).

Sabiendo que el factor C mide estabilidad emocional, de manera que a mayor puntuación en el factor, mayor nivel de estabilidad emocional presenta el sujeto, los resultados anteriores nos dicen que, cuando el punto medio en la escala de respuesta es “término medio” los sujetos que tienen más probabilidad de escoger esta opción como respuesta son los que presentan mayor nivel de estabilidad emocional. Sin embargo, cuando el punto medio en la escala de respuesta es “no estoy seguro” o “?”, los sujetos que tienen más probabilidad de escoger esta opción como respuesta son los que presentan menor nivel de estabilidad.

En resumen, tanto los sujetos más estables emocionalmente, como los menos estables, tienden a responder situándose en el punto medio de un continuo, sólo que los menos estables lo hacen cuando la escala de respuesta contiene la categoría de intermedia “no estoy seguro” o “?”, mientras que los más estables lo hacen cuando la escala de respuesta contiene la categoría “término medio”.

La cuestión es averiguar por qué los individuos menos estables tienen mayor probabilidad de escoger NES o ?, y por qué los más estables tienen mayor probabilidad de escoger el punto medio TM. Quizá los individuos menos estables emocionalmente manifiestan las características de su personalidad eligiendo unas categorías de respuesta que implican inseguridad. En el segundo caso, para averiguar por qué los individuos más

estables tienen más probabilidad de escoger el punto medio TM quizá habría que analizar el contenido de los ítems de dicho factor.

A la vista de los análisis realizados, el estudio de las razones que se ha llevado a cabo no ha ayudado a comprender el funcionamiento de las alternativas de respuesta para el factor C, en comparación con el funcionamiento de las alternativas para los factores H y Q4.

Sin embargo, hay que destacar que, a pesar de que los tres puntos medios presentan un alto porcentaje de veces que se elige la razón 4 (me siento en el punto medio), también es cierto que el resto de las razones que se aportan para las tres alternativas centrales son razones que no tienen nada que ver con sentirse en el punto medio de la cuestión que se plantea. Es decir, más de un 40% de las veces que se elige alguno de los tres puntos medios del estudio, se escoge esta categoría de respuesta por otra razón que no es estar en el punto medio.

Esto plantea hasta qué punto vale la pena utilizar esos puntos medios o alternativas centrales si gran parte de las veces el significado que se le otorga a estas categorías de respuesta no es el punto medio. La cuestión sería ver si realmente se gana algo presentando un formato de respuesta de tres alternativas frente a un formato de respuesta dicotómico. Para intentar responder a esta cuestión se ha realizado un estudio de la validez criterial de los tres factores con las tres escalas de respuesta politómicas y con la escala de respuesta dicotómica (ver tabla 5.5).

Tabla 5.5. Correlaciones de Pearson entre los factores C, H y Q4 del 16 PF con sus cuatro formatos de respuesta, y los criterios autoestima y ansiedad-rasgo (DIC: escala de respuesta dicotómica; NES: con punto medio “no estoy seguro”; TM: con punto medio “término medio”; QM: con punto medio “?”)

		Autoestima	Ansiedad-Rasgo
FACTOR C	DIC	-,4107	,3263
	NES	,3789	-,3438
	QM	,2843	-,5072
	TM	,2545	-,4869
FACTOR H	DIC	-,3248	,2582
	NES	,2365	-,3732
	QM	,2037	-,3457
	TM	,2293	-,3552
FACTOR Q4	DIC	,3308	-,3469
	NES	-,2289	,5053
	QM	-,2113	,4940
	TM	-,2227	,4882

*** Todas las correlaciones fueron significativas con $p < 0.01$**

Como se puede observar en la tabla 5.5, las escalas de respuesta dicotómicas presentan unas correlaciones significativas con los criterios en el sentido opuesto al esperado, mientras que el resto de las escalas de respuesta presenta correlaciones significativas y en el signo esperado. La estabilidad emocional (C+) se espera que correlacione positivamente con la autoestima y negativamente con la ansiedad-rasgo, el atrevimiento (H+) debería presentar correlaciones positivas con la autoestima y negativas con la ansiedad-rasgo, y la tensión (Q-) debería mostrar correlaciones negativas

con autoestima y positivas con ansiedad-rasgo, ya que la tensión es una dimensión de la ansiedad.

Este patrón de correlaciones no se cumple para ningún factor con escala dicotómica, mientras que se cumple para los tres factores con los tres tipos de escala de respuesta (con NES, con TM y con ?).

Así pues, las correlaciones con los criterios no muestran precisamente que los items con escala de respuesta dicotómica presenten validez criterial, puesto que todas las correlaciones son significativas y en sentido esperado. Parece por tanto que es más razonable mantener escalas de respuesta de tres puntos antes que decidirse por escalas de respuesta dicotómicas.

CAPÍTULO 6.

DISCUSIÓN

El objetivo del proceso de medición en Psicología es representar las relaciones observadas entre las propiedades de los objetos mediante el sistema numérico adecuado que preserve dichas relaciones. La representación cuantitativa de estas características permite establecer comparaciones entre sujetos y distintos análisis matemáticos empleando las técnicas estadísticas apropiadas. Cuando una conducta es abierta se puede observar directamente, y se registra asignándole una puntuación en una escala; si la conducta es encubierta, es decir, si no es directamente observable, se administra un test o un cuestionario y se obtendrá así una serie de respuestas que se transformarán en una puntuación numérica.

Los tests pretenden ser medidas objetivas y precisas de un constructo teórico. Estas medidas se obtienen a partir de una muestra representativa de conductas consideradas como indicadores de dicho constructo en situaciones estandarizadas. Son además una técnica que permite recoger la información de manera sistemática, es decir, se siguen

siempre las mismas instrucciones y los sujetos responden a las mismas cuestiones, con lo que se evita así que las diferencias entre las puntuaciones obtenidas se deban a diferencias en la aplicación del test. Finalmente, la puntuación obtenida por los sujetos en el test se compara con la puntuación que presenta un grupo normativo en el rasgo medido o con la puntuación en algún criterio externo y se lleva a cabo una serie de predicciones o inferencias acerca de otras conductas.

Los tests y demás instrumentos de medición forman parte del quehacer cotidiano del psicólogo, ya que se emplean con asiduidad en ámbitos como la educación, la selección de personal, la orientación profesional, etc. Los resultados obtenidos a partir de dichas mediciones suelen tener bastante repercusión en la vida de las personas y por ello, si se utilizan incorrectamente, los tests pueden ser instrumentos de medida que lleven al psicólogo a graves errores a la hora de la toma de decisiones. Aunque en todas las ciencias el conocimiento de los instrumentos de medida es fundamental, el psicólogo debe hacer un esfuerzo mayor en este sentido debido a que los atributos psicológicos no se pueden medir directamente, y sin embargo se utilizan para explicar la conducta humana.

Al obtener de este modo una puntuación del sujeto en el rasgo, se está asumiendo que la puntuación obtenida representa el valor auténtico en el rasgo que presenta el sujeto, es decir, se considera un estimador muestral de dicho valor (Martínez Arias, 1995). Sin embargo, como no se puede saber si la puntuación observada es la puntuación verdadera en el rasgo porque siempre está presente la influencia del azar, el proceso de medición en Psicología es en realidad un proceso de inferencia.

Los valores observados son los que han de preservar las relaciones del sistema empírico al que representan, y cuando no lo hacen vulneran el isomorfismo que ha de existir entre los sistemas empírico y numérico. En la medida que la puntuación observada no represente fielmente a la puntuación verdadera del sujeto en el rasgo, se verán afectadas la fiabilidad y la validez del instrumento de medida empleado.

En este sentido, el sistema de puntuación de las escalas de respuesta ordinales tiene una importancia fundamental ya que, si no preserva el isomorfismo que ha de existir entre los sistemas relacionales empírico y numérico y se verán afectadas la fiabilidad y la validez del instrumento de medida empleado.

Pero la hora de proponer un sistema de puntuación también es importante el significado que se le atribuye las alternativas de respuesta, y concretamente al punto medio, ya que la puntuación que se le otorgue a esta categoría de respuesta debería depender de dicho significado. Normalmente se asume que el significado que los sujetos atribuyen a esta categoría es el mismo que le atribuye el investigador, pero no siempre es así (Wang y Stanley, 1970).

Se ha escogido el 16 PF como instrumento de medida objeto de estudio, ya que para todos sus items presenta escalas de respuesta tipo Likert. Además, se trata de un instrumento ampliamente utilizado tanto en el campo profesional como en la investigación, con lo que un sistema de puntuación que no sea congruente con el funcionamiento de las categorías

de la escala de respuesta tendría una amplia repercusión, ya que se vería afectada la fiabilidad y la validez del instrumento.

Los análisis realizados para comprobar si las categorías de respuesta se presentaban ordenadas aplicando el modelo nominal de Bock mostraron un ajuste aceptable para cada factor. Respecto a los parámetros a estimados, los de los items de los factores H y Q4 (con punto medio = “término medio” en la escala de respuesta) son los únicos que se presentan ordenados tal y como se asume en un principio para las categorías de respuesta (“1. No”, “2. Término medio”, “3. Sí”).

Para los items de los restantes factores (factores C, E, F, G, L y O) los parámetros a estimados no se presentan ordenados. Concretamente, para los items de los factores C y F, cuya escala de respuesta incluye la categoría “no estoy seguro” como punto medio, el parámetro a estimado para la categoría intermedia presenta el menor valor, seguido del parámetro a estimado para la categoría de respuesta 1 y, por último, del parámetro a estimado para la categoría de respuesta 3 (orden 213; NES,NO,SI), y para los items de los factores L, G y O, cuya escala de respuesta incluye la categoría “término medio” como punto medio, el parámetro a estimado para la categoría intermedia presenta el menor valor, seguido del parámetro a estimado para la categoría de respuesta 1 y, por último, del parámetro a estimado para la categoría de respuesta 3 (orden 213; TM,NO,SI). Según estos resultados, los sujetos con menor valor en el rasgo serían los que presentarían una mayor probabilidad de escoger la categoría intermedia como respuesta, mientras que los sujetos que tienen más probabilidad de responder la categoría 1 serían los que presentan valores medios en el rasgo.

Respecto a los ítems del factor E, los parámetros a estimados presentan una ordenación diferente para distintos ítems, y en ningún caso este orden es el que se asume en el sistema de puntuación del 16 PF. Para el ítem 138 el orden de los parámetros a correspondientes a las categorías de respuesta es 312 (SI,NO,TM), de modo que ninguna categoría de respuesta se presenta en el lugar que le correspondería a priori. Para el resto de los ítems el parámetro a estimado para la categoría de respuesta 1 presenta el menor valor, seguido del parámetro a estimado para la categoría de respuesta 3 y, por último, del parámetro a estimado para el punto medio (NO,SI,TM).

Por lo que respecta al segundo estudio, realizado con la segunda muestra, se planteó con diversos objetivos. Por una parte, se quería comprobar si los puntos medios analizados con anterioridad presentaban el mismo funcionamiento en una nueva muestra; además, se incluyó el estudio del punto medio que aparece en todos los ítems de la nueva versión del 16 PF (“?”) para comprobar si el funcionamiento del mismo es el que asume el sistema de puntuación del cuestionario; por último, se quería comprobar si los ítems analizados en el estudio primero se comportaban de una manera diferente en función de la escala de respuesta, es decir, se planteó estudiar si el funcionamiento de un ítem podía ser diferente en el caso de que se presentara con escalas de respuestas diferentes. Además, se planteó estudiar el significado que tiene el punto medio para los sujetos, para ver si los resultados obtenidos a partir de este análisis podían explicar el hecho de que algunas categorías de respuesta no se presentaran ordenadas.

Para llevar a cabo estos objetivos, se elaboró una batería en la que se incluyeron los factores C y Q4 del estudio anterior, y una serie de items del factor H correspondientes al 16 PF-5, además de una lista de posibles razones por las que se podía elegir el punto medio.

Los análisis realizados para comprobar si las categorías de respuesta se presentaban ordenadas aplicando el modelo nominal de Bock a los factores con escalas politómicas mostraron un ajuste aceptable para cada factor, y el modelo de dos parámetros empleado para estimar la bondad de ajuste de los datos dicotómicos también ofreció resultados satisfactorios para todos los items de los factores.

Respecto al orden de las categorías de respuesta, los parámetros a estimados para los items de los factores H y Q4 se presentan ordenados para todas las categorías de respuesta (orden 123; NO, TM, SI). Sin embargo, para los items del factor C los parámetros a presentan un orden distinto según la escala de respuesta: cuando la escala de respuesta tiene como punto medio la interrogación (“?”) y cuando el punto medio es “2. NES”, los parámetros a estimados para los items del factor C se presentan en el orden 213 (NES/?, NO, SI). Según estos resultados, los sujetos que tienen más probabilidad de responder el punto medio de la escala de respuesta (“?” o “2. NES”) son aquéllos que presentan un valor más bajo en el rasgo, mientras que los sujetos que tienen más probabilidad de responder la categoría de respuesta 1 son los que presentan valores intermedios en el rasgo.

Según esto, el significado que debería tener el punto medio para los sujetos, cuando responden los items de los factores H y Q4, habría de ser mayoritariamente “estar en el punto medio de un continuo”, mientras que para los items del factor C el significado otorgado a la categoría intermedia debería ser otro. Sin embargo, el porcentaje de veces que se aporta la razón 4 (me siento en el punto medio) es similar para los tres factores respecto a cada escala de respuesta.

A pesar de ello, hay que destacar que el resto de las razones que se aportan para las tres alternativas centrales son razones que no tienen nada que ver con sentirse en el punto medio de la cuestión que se plantea. Es decir, más de un 40% de las veces que se elige alguno de los tres puntos medios del estudio, se escoge esta categoría de respuesta por otra razón que no es estar en el punto medio.

Esto plantea hasta qué punto vale la pena utilizar esos puntos medios o alternativas centrales si gran parte de las veces el significado que se le otorga a estas categorías de respuesta no es el punto medio. La cuestión sería ver si realmente se gana algo presentando un formato de respuesta de tres alternativas frente a un formato de respuesta dicotómico. Para intentar responder a esta cuestión se realizó un estudio de la validez criterial de los tres factores con las tres escalas de respuesta politómicas y con la escala de respuesta dicotómica, pudiéndose comprobar que las escalas de respuesta dicotómica presentan correlaciones significativas pero de signo contrario al esperado con ambos criterios (autoestima y ansiedad-rasgo), mientras que las escalas politómicas presentan correlaciones significativas y del signo

esperado. Este resultado no apoya por tanto la validez de las escalas de respuesta dicotómica.

Otra cuestión es averiguar por qué los individuos menos estables tienen mayor probabilidad de escoger NES o ?, y por qué los más estables tienen mayor probabilidad de escoger el punto medio TM. Quizá los individuos menos estables emocionalmente manifiestan las características de su personalidad eligiendo unas categorías de respuesta que implican inseguridad. En el segundo caso, para averiguar por qué los individuos más estables tienen más probabilidad de escoger el punto medio TM quizá habría que analizar el contenido de los items de dicho factor.

Por otra parte, de los dos estudios llevados a cabo, cada uno con una muestra, destaca también que en el análisis factorial confirmatorio llevado a cabo en el primer estudio hay bastantes factores que no sostienen la unidimensionalidad de los items que los integran.

En el primer estudio, realizado con la primera muestra, los análisis factoriales llevados a cabo para comprobar la unidimensionalidad de los factores, mostraron que los items de algunos de los factores de este cuestionario (factores F (subgrupo con punto medio = “término medio”), M, N, Q2 y Q3 (ambos subgrupos)) presentaban problemas para sostener la unidimensionalidad de los mismos, mostrando unos resultados en los análisis confirmatorios realizados que indicaban que dichos items no medían una dimensión común. El resto de los factores analizados (factores C, E, F (subgrupo con punto medio = “no estoy seguro”), G, H, L, O y Q4) sí presentaron resultados satisfactorios en los análisis factoriales.

En el segundo estudio, los resultados obtenidos tras realizar el análisis factorial de cada factor indicaron que un modelo unifactorial presentaba un ajuste aceptable para todas las escalas de respuesta del factor Q4 y del factor C (“?”, “TM”, “NES” y *escala de respuesta dicotómica*). Sin embargo para el factor H el ajuste no era satisfactorio y hubo que eliminar el ítem 137, que era el que presentaba la saturación factorial más baja, tras lo cual los índices de ajuste del modelo mejoraron.

Los resultados obtenidos con el análisis factorial muestran bastantes semejanzas entre los factores presentes en los dos estudios. Si bien en la muestra del primer estudio había factores del 16 PF que no mostraban unidimensionalidad, mientras que con la muestra del segundo estudio no hubo problemas con ningún factor, excepto con un ítem del factor H, en el segundo estudio se analizan únicamente tres de los factores que aparecían en el primer estudio, los cuales además ya habían mostrado resultados satisfactorios en la primera muestra.

En resumen, habría que seguir estudiando el funcionamiento de las escalas de respuesta del 16 PF en muestras diferentes, para comprobar si se mantiene el orden que presentan las categorías de las escalas de respuesta, además de seguir analizando el significado que los sujetos atribuyen al punto medio. Así mismo, habría que estudiar la estructura factorial del cuestionario a fin de comprobar si los factores que presentan problemas en el primer estudio, y que no fueron incluidos en el segundo estudio, siguen comportándose de la misma manera en muestras distintas.

Referencias bibliográficas

- Aftanas, M. S. (1988). Theories, models and standards systems of measurement. *Applied Psychological measurement*, 12, (4), 325-338.
- Aiken, L. R. (1996). *Rating scales and checklists: evaluating behavior, personality and attitudes*. New York: John Wiley & Sons, Inc.
- Allen, M. J. y Yen, W. M. (1979). *Introduction to measurement theory*. Belmont, California: Wadsworth, Inc.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, (4), 561-573.
- Andrich, D. (1982). An extension of the Rasch model for ratings providing both location and dispersion parameters. *Psychometrika*, 47, 105-113.
- Andrich, D. y Schoubroeck, L. (1989). The General Health Questionnaire: a psychometric analysis using latent trait theory. *Psychological Medicine*, 19, 469-485.
- Birnbaum, A. (1968). En Lord, F. M. y Novick, m. R.: *Statistical theories of mental test scores*. Reading Mass., Addison-Wesley.
- Bock, R. D. (1972). Estimating item parameters and latent ability when the responses are scored in two or more nominal categories. *Psychometrika*, 37, 29-51.
- Bock, R. D. (1975). *Multivariate Statistical Methods in Behavioral Research*. New York: McGraw-Hill.

- Bock, R. D. (1997). The nominal categories model. En W. J. van der Linden and R. K. Hambleton (eds.): *Handbook of modern item response theory*. New York Inc.: Springer-Verlag.
- Bock, R. D. y Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46, (4), 443-459.
- Bolton, B. (1985). Discriminant analysis of Holland's occupational types using the 16PF. *Journal of Vocational Behavior*, 27, 210-217.
- Campbell, N. R. (1928). *An account of the principles of measurement and calculation*. London: Longmans, Green.
- Cattell R. B. y Krug, S. (1986). The number of factors in the 16PF: A review of the evidence with special emphasis on methodological problems. *Educational and Psychological Measurement*, 46, 509-522.
- Cattell R. B., Cattell, A. K. S., Cattell H. E. P. (1993). *Handbook for the Sixteen Personality Factor Questionnaire*. Champaign, IL: IPAT.
- Cattell, R. B. ; Eber, H. W. y Delhees, K. H. (1968). A large sample cross validation of the personality trait structure of the 16PH with some clinical implications. *Multivariate Behavioral Research, special issue*, 107-132.
- Cattell, R. B. ; Eber, H. W. y Tatsuoka, M. M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire*. Champaign, IL: IPAT.
- Cheung, K. C. y Mooi, L. C. (1994). A comparison between the rating scale model and dual scaling for Likert scales. *Applied Psychological Measurement*, 18, (1), 1-13.
- Coombs, C. H.; Dawes, R. M. y Tversky, A. (1981). *Introducción a la psicología matemática*. Madrid: Alianza Editorial, S. A.

- Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement*, 6, 475-494.
- Drasgow, F.; Levine, M. V.; Tsien, S.; Williams, B. y Mead, A. D. (1995). Fitting polytomous item response theory models to multiple-choice tests. *Applied Psychological Measurement*, 19, (2), 143-165.
- Dubois, B. y Burns, J. A. (1975). An analysis of the meaning of the question mark response category in attitude scales. *Educational and Psychological Measurement*, 35, 869-884.
- Edwards, A. L. (1946). A critique of "neutral" items in attitude scales constructed by the method of equal appearing intervals. *Psychological Review*, 53, 159-169.
- Goldberg, D. P. (1978). *Manual of the General Health Questionnaire*. Windsor: National Foundation for Educational Research.
- Goldberg, G. (1971). Response format in attitude scales. Unpublished manuscript, Northwestern University.
- Goodchild, M. E. y Duncan-Jones, P. (1985). Chronicity and the GHQ. *British Journal of Psychiatry*, 146, 55-61.
- Gray-Little, B., Williams, V. S. L. Hancock, T. D. (1997). An items response theory analysis of the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, 23, (5), 443-451.
- Hambleton, R. K., Swaminathan, H. y Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park: SAGE.
- Hanisch, K. (1992). The Job Descriptive Index Revisited: Questions about the question mark. *Journal of applied Psychology*, 77, 3, 377-382.
- Hathaway, S. R. y McKinley, J. C. (1967). *Minnesota Multiphasic Personality Inventory. Manual Revised*. New York: The Psychological Corporation.

- Hölder, O. (1901). Die Axiome de Quantität die Lehre von Mass. *Berichte ueber die Verhandlugen der Königlich Sachsische Gessellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Class*, 53, 1-64.
- Howarth, E. y Browne, J. A. (1971). An item-factor-analysis of the 16PF. *Personality*, 2, 117-139.
- Hulin, C. L. y Mayer, L. J. (1979). Psychometric equivalence of a translation of the Job Descriptive Index into hebrew. *Journal of Applied Psychology*, 71, (1), 83-94.
- Jáñez, L. (1989). *Fundamentos de Psicología Matemática*. Madrid: Pirámide.
- Joreskog, K. G. y Sörbom, D. (1993). *LISREL 8 user's reference guide*. Chicago, IL: Scientific Software International, Inc.
- Kaplan, K. J. (1972). On the ambivalence-indifference problem in attitude theory: A suggested modification of the semantic differential technique. *Psychological Bulletin*, 77, 361-372.
- Krug, S. E. y Johns, E. F. (1990). The 16 Personality Factor Questionnaire. In Watkins, C. E. y Campbell, V. L. (eds.): *Testing in counseling practice*. Hillsdale, New Jersey: Lawrence Erlbaum associates, Publishers.
- Lord, F. M. (1953). The relation of test score to the trait underlying the test. *Educational and Psychological Measurement*, 13, 517-548.
- Lord, F. M. y Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, Massachussets: Addison-Wesley.
- Martínez Arias, R. (1995). *Psicometría: teoría de los tests psicológicos y educativos*. Madrid: Síntesis.

- Maydeu-Olivares, A.; Drasgow, F. y Mead, A. D. (1994). Distinguishing among parametric item response models for polichotomous ordered data. *Applied Psychological Measurement*, 18, (3), 245-256.
- Meddis, R. (1972). Bipolar factors in mood adjective checklists. *British Journal of Social and clinical Psychology*, 11, 283-284.
- Mellenberg, G. J. (1995). Conceptual notes on models for discrete polytomous item responses. *Applied Psychological Measurement*, 19, (1), 91-100.
- Muñiz, J. (1992). *Teoría clásica de los tests*. Madrid: Pirámide.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, N°. 17.
- Samejima, F. (1972). A general model for free-response data. *Psychometrika Monograph*, No. 18.
- Santisteban, C. (1990). *Psicometría. Teoría y práctica en la construcción de tests*. Madrid: Norma.
- Saville, P. y Blinkhorn, S. (1981). Reliability, homogeneity and the construct validity of Cattell's 16PF. *Personality and Individual differences*, 2, 325-333.
- Seisdedos, N. (1988). *STAI. Cuestionario de Ansiedad Estado-Rasgo. Manual*. Madrid. TEA Ediciones.
- Seisdedos, N. (1989). *16 PF, Cuestionario factorial de personalidad (adolescentes y adultos)*. Madrid: TEA Ediciones, S. A. 9ª edición.
- Smith, P. C., Kendall, L. M. y Hulin, C. L. (1969). *The measurement properties of satisfaction in work and retirement: A strategy for the study of attitudes*. Skokie, IL: Rand-McNally.

- Spielberger, C. D.; Gorsuch, R. L. y Lushene, R. E. (1970). *STAI. manual for the State-Trait Anxiety Inventory (Self-Evaluation Questionnaire)*. Palo Alto, CA: Consulting Psychologists Press.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 677-680.
- Stevens, S. S. (1951). Mathematics, measurement and psychophysics. En S. S. Stevens (ed.): *Handbook of Experimental Psychology*. Nueva York: Wiley.
- Suppes, P. y Zinnes, J. L. (1963). Basic measurement theory. En R. D. Luce, R. R. Bush y E. Galanter (Eds.): *Handbook of mathematical psychology*. New York: Wiley.
- Thissen, D. (1988). *MULTILOG: Multiple, categorical item analysis and test scoring using item response theory (Version 5.1)*. Computer program. Mooresville IN: Scientific Software.
- Thissen, D. y Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika*, 51, (4), 567-577.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34, 273-286.
- Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.
- Walter, V. (1985). *Personal Career Development Profile manual*. Champaign, IL: IPAT.
- Wang, M. D. y Stanley, J. C. (1970). Differential weighting: a review of methods and empirical studies. *Review of Educational Research*, 40, 663-705.

- Wherry, R. J. (1982). The control of bias in ratings: A theory of ratings. (Edited and Comments by C. J. Bartlett). *Personnel Psychology*, 35, 521-551.
- Wiggins J. S. y Pincus, A. L. (1992). Conceptions of personality disorders and dimensions of personality. *Psychological Assessment: J. Consult. Clin. Psychology*, 1, 305-316.
- Worthy, M. (1969). Note on scoring midpoint responses in extreme response style scores. *Psychological Reports*, 24, 189-190.

ANEXO I



**Unidad de Investigación de Modelos de Ecuaciones Estructurales y Teoría de Respuesta al Ítem
(UIMETRI)
Universitat de València**

Estimado/a colaborador/a:

En las páginas siguientes encontrarás una serie de cuestiones que permitirán conocer tus actitudes e intereses. **Las respuestas a estas cuestiones sólo se utilizarán con fines de investigación científica. Se garantiza el anonimato de tus contestaciones. Por ello, te rogamos que contestes con toda sinceridad y confianza todas las preguntas que te formulamos.** Gracias a colaboraciones desinteresadas como la tuya podremos aumentar nuestro conocimiento sobre cómo mejorar la medición de los constructos en Psicología, permitiendo el avance de esta ciencia y la realización de intervenciones adecuadas.

Muchas gracias por tu colaboración

INSTRUCCIONES

A continuación encontrarás una serie de cuestiones que permitirán conocer tus actitudes e intereses. En general, no existen contestaciones correctas o incorrectas, porque las personas tienen distintos intereses y ven las cosas desde distintos puntos de vista. Las cuestiones que se formulan presentan dos o tres alternativas de respuesta, y cada una de ellas va precedida de un número. **Para indicar tu contestación debes RODEAR CON UN CÍRCULO la opción elegida, e ignorar la línea que aparece junto a cada cuestión.** Por ejemplo:

1. Me gusta presenciar una competición deportiva.
 1. Verdadero 2. Término medio 3. Falso _____
2. Prefiero a las personas reservadas.
 1. Sí 2. No _____
3. El dinero no hace la felicidad.
 1. Verdadero 2. ? 3. Falso _____

Verás que la mayoría de las cuestiones aparecen varias veces con pequeñas modificaciones. Eso se debe a las características del estudio. Contesta cada cuestión sin pensar en lo que has contestado anteriormente.

Al contestar ten en cuenta lo siguiente:

- No pienses demasiado el contenido de las frases ni emplees mucho tiempo en decidirte. Las frases son muy cortas para dar todos los detalles que quisieras; en el ejemplo se ha puesto "presenciar una competición deportiva", y tal vez a ti te guste más el fútbol que el baloncesto; debes contestar pensando en lo que es habitual para ti.
- Contesta sinceramente. No señales tus respuestas pensando en lo que "es bueno" o lo que "interesa" para impresionar al examinador.
- Procura no dejar ninguna cuestión sin contestar. Es posible que alguna no tenga nada que ver contigo (porque no se aplica perfectamente a tu caso); intenta elegir la respuesta que vaya mejor con tu modo de ser. Tal vez algunas frases te parezcan muy personales; no te preocupes y recuerda que serán tratadas confidencialmente. Por otra parte, al obtener los resultados no se consideran las respuestas una a una, sino globalmente.

Una vez hayas respondido a las cuestiones de una página, **NO VUELVAS ATRÁS A MENOS QUE SE TE INDIQUE.**

1. Me pongo algo nervioso ante animales salvajes, incluso cuando están encerrados en fuertes jaulas.
1. Sí 2. No estoy seguro 3. No _____
2. Algunas veces me "sacan de quicio" de un modo insoportable pequeñas cosas, aunque reconozca que son triviales.
1. Sí 2. Término medio 3. No _____
3. Si cometo una falta social desagradable puedo olvidarla pronto.
1. Sí 2. No estoy seguro 3. No _____
4. En ocasiones dudo si la gente con quien estoy hablando se interesa realmente por lo que digo.
1. Sí 2. Término medio 3. No _____
5. Suelo enfadarme con las personas demasiado pronto.
1. Sí 2. Término medio 3. No _____
6. En mi vida personal consigo casi siempre todos mis propósitos.
1. Verdadero 2. No estoy seguro 3. Falso _____
7. En ocasiones, contrariedades muy pequeñas me irritan mucho.
1. Sí 2. Término medio 3. No _____
8. Me suelo sentir desconcertado si de pronto paso a ser el centro de la atención en un grupo social.
1. Verdadero 2. ? 3. Falso _____
9. Me considero una persona socialmente muy atrevida y comunicativa.
1. Verdadero 2. ? 3. Falso _____
10. Me cuesta bastante hablar delante de un grupo numeroso de personas.
1. Verdadero 2. ? 3. Falso _____
11. Algunas personas parecen ignorarme o evitarme, aunque no sé por qué.
1. Verdadero 2. No estoy seguro 3. Falso _____
12. Cuando estoy con un grupo, normalmente me siento, escucho y dejo que los demás lleven el peso de la conversación.
1. Verdadero 2. ? 3. Falso _____
13. Cuando algo realmente me pone furioso, suelo calmarme muy pronto.
1. Sí 2. Término medio 3. No _____
14. Cuando me uno a un nuevo grupo, normalmente encajo pronto.
1. Verdadero 2. ? 3. Falso _____
15. A veces no puedo dormir porque tengo una idea que me da vueltas en la cabeza.
1. Verdadero 2. No estoy seguro 3. Falso _____

Escribe tu edad y **RODEA CON UN CÍRCULO** la opción que se adecúe a tu caso.

1. Indica tu edad: _____ años

2. Señala los estudios de mayor nivel que has completado:

1. Sabes leer y escribir.
2. Primarios (Certificado de Escolaridad, Graduado Escolar).
3. Formación Profesional-1er. grado.
4. Formación Profesional-2º. grado, B.U.P. o Bachiller.
5. Titulación Universitaria Media (Escuelas Técnicas, Profesores de E.G.B., Graduados Sociales, ATS-DUE, etc.).
6. Licenciados, Técnicos Superiores y Doctores.

3. Indica tu sexo:

1. Varón
2. Mujer

4. Señala cuál es tu ocupación principal:

1. Estudiante
2. Trabajador/a en activo
3. Desempleado/a
4. Ama de casa

5. Excluyendo los libros de texto, indica cuántos libros lees al cabo de un año (novela, poesía, etc.): _____

6. Señala cuál es la lengua en la que te expresas habitualmente:

1. Valenciano
2. Castellano

A continuación, se presenta una nueva serie de items con dos o tres alternativas de respuesta. Para indicar tu contestación **RODEA CON UN CÍRCULO** la opción elegida.

16. En ocasiones dudo si la gente con quien estoy hablando se interesa realmente por lo que digo.
 1. Sí 2. No estoy seguro 3. No _____
17. En mi vida personal consigo casi siempre todos mis propósitos.
 1. Verdadero 0. Falso _____
18. Suelo enfadarme con las personas demasiado pronto.
 1. Sí 2. No estoy seguro 3. No _____
19. Cuando algo realmente me pone furioso, suelo calmarme muy pronto.
 1. Sí 2. No estoy seguro 3. No _____
20. Cuando estoy con un grupo, normalmente me siento, escucho y dejo que los demás lleven el peso de la conversación.
 1. Verdadero 2. Término medio 3. Falso _____
21. Algunas personas parecen ignorarme o evitarme, aunque no sé por qué.
 1. Verdadero 0. Falso _____
22. Cuando me uno a un nuevo grupo, normalmente encajo pronto.
 1. Verdadero 2. Término medio 3. Falso _____
23. Si cometo una falta social desagradable puedo olvidarla pronto.
 1. Sí 0. No _____
24. A veces no puedo dormirme porque tengo una idea que me da vueltas en la cabeza.
 1. Verdadero 0. Falso _____
25. Algunas veces me "sacan de quicio" de un modo insoportable pequeñas cosas, aunque reconozca que son triviales.
 1. Sí 2. No estoy seguro 3. No _____
26. En ocasiones, contrariedades muy pequeñas me irritan mucho.
 1. Sí 2. No estoy seguro 3. No _____
27. Me suelo sentir desconcertado si de pronto paso a ser el centro de la atención en un grupo social.
 1. Verdadero 2. Término medio 3. Falso _____
28. Me cuesta bastante hablar delante de un grupo numeroso de personas.
 1. Verdadero 2. Término medio 3. Falso _____
29. Me pongo algo nervioso ante animales salvajes, incluso cuando están encerrados en fuertes jaulas.
 1. Sí 0. No _____
30. Me considero una persona socialmente muy atrevida y comunicativa.
 1. Verdadero 2. Término medio 3. Falso _____

Los siguientes items presentan cuatro alternativas de respuesta. Para indicar tu contestación debes **RODEAR CON UN CÍRCULO** la opción elegida.

1. Siento que soy una persona digna de estima, al menos en igual medida que los demás.

1. Estoy muy de acuerdo
2. Estoy de acuerdo
3. No estoy de acuerdo
4. Estoy muy en desacuerdo

2. Creo tener varias cualidades buenas.

1. Estoy muy de acuerdo
2. Estoy de acuerdo
3. No estoy de acuerdo
4. Estoy muy en desacuerdo

3. Me inclino a pensar que, en conjunto, soy un fracaso.

1. Estoy muy de acuerdo
2. Estoy de acuerdo
3. No estoy de acuerdo
4. Estoy muy en desacuerdo

4. Puedo hacer las cosas tan bien como la mayoría de las otras personas

1. Estoy muy de acuerdo
2. Estoy de acuerdo
3. No estoy de acuerdo
4. Estoy muy en desacuerdo

5. Creo que no tengo muchos motivos para enorgullecirme.

1. Estoy muy de acuerdo
2. Estoy de acuerdo
3. No estoy de acuerdo
4. Estoy muy en desacuerdo

6. Asumo una actitud positiva hacia mí mismo.

1. Estoy muy de acuerdo
2. Estoy de acuerdo
3. No estoy de acuerdo
4. Estoy muy en desacuerdo

7. En general, estoy satisfecho conmigo mismo.

1. Estoy muy de acuerdo
2. Estoy de acuerdo
3. No estoy de acuerdo
4. Estoy muy en desacuerdo

8. Desearía sentir más aprecio por mí mismo.

1. Estoy muy de acuerdo
2. Estoy de acuerdo
3. No estoy de acuerdo
4. Estoy muy en desacuerdo

9. A veces me siento realmente inútil.

1. Estoy muy de acuerdo
2. Estoy de acuerdo
3. No estoy de acuerdo
4. Estoy muy en desacuerdo

10. A veces pienso que no sirvo para nada.

1. Estoy muy de acuerdo
2. Estoy de acuerdo
3. No estoy de acuerdo
4. Estoy muy en desacuerdo

A continuación encontrarás una frases que se utilizan corrientemente para describirse a uno mismo. Lee cada frase y **RODEA CON UN CÍRCULO** la opción que indique mejor **CÓMO TE SIENTES EN GENERAL** en la mayoría de las ocasiones. No emplees demasiado tiempo en cada frase.

1. Me siento bien.
 1. Casi nunca
 2. A veces
 3. A menudo
 4. Casi siempre
2. Me canso rápidamente.
 1. Casi nunca
 2. A veces
 3. A menudo
 4. Casi siempre
3. Siento ganas de llorar.
 1. Casi nunca
 2. A veces
 3. A menudo
 4. Casi siempre
4. Me gustaría ser tan feliz como otros.
 1. Casi nunca
 2. A veces
 3. A menudo
 4. Casi siempre
5. Pierdo oportunidades por no decidirme pronto.
 1. Casi nunca
 2. A veces
 3. A menudo
 4. Casi siempre
6. Me siento descansado.
 1. Casi nunca
 2. A veces
 3. A menudo
 4. Casi siempre
7. Soy una persona tranquila, serena y sosegada.
 1. Casi nunca
 2. A veces
 3. A menudo
 4. Casi siempre
8. Veo que las dificultades se amontonan y no puedo con ellas.
 1. Casi nunca
 2. A veces
 3. A menudo
 4. Casi siempre
9. Me preocupo demasiado por cosas sin importancia.
 1. Casi nunca
 2. A veces
 3. A menudo
 4. Casi siempre
10. Soy feliz.
 1. Casi nunca
 2. A veces
 3. A menudo
 4. Casi siempre
11. Suelo tomar las cosas demasiado seriamente.
 1. Casi nunca
 2. A veces
 3. A menudo
 4. Casi siempre
12. Me falta confianza en mí mismo.
 1. Casi nunca
 2. A veces
 3. A menudo
 4. Casi siempre
13. Me siento seguro.
 1. Casi nunca
 2. A veces
 3. A menudo
 4. Casi siempre
14. Evito enfrentarme a las crisis o dificultades.
 1. Casi nunca
 2. A veces
 3. A menudo
 4. Casi siempre
15. Me siento triste (melancólico).
 1. Casi nunca
 2. A veces
 3. A menudo
 4. Casi siempre
16. Estoy satisfecho.
 1. Casi nunca
 2. A veces
 3. A menudo
 4. Casi siempre
17. Me rondan y molestan pensamientos sin importancia.
 1. Casi nunca
 2. A veces
 3. A menudo
 4. Casi siempre
18. Me afectan tanto los desengaños, que no puedo olvidarlos.
 1. Casi nunca
 2. A veces
 3. A menudo
 4. Casi siempre
19. Soy una persona estable.
 1. Casi nunca
 2. A veces
 3. A menudo
 4. Casi siempre
20. Cuando pienso sobre asuntos y preocupaciones actuales, me pongo tenso y agitado.
 1. Casi nunca
 2. A veces
 3. A menudo
 4. Casi siempre

A continuación presentamos una última serie de cuestiones con dos o tres alternativas de respuesta. **RODEA CON UN CÍRCULO** la opción elegida.

46. Me suelo sentir desconcertado si de pronto paso a ser el centro de la atención en un grupo social.
1. Verdadero 2. No estoy seguro 3. Falso _____
47. En ocasiones, contrariedades muy pequeñas me irritan mucho.
1. Sí 0. No _____
48. Suelo enfadarme con las personas demasiado pronto.
1. Sí 0. No _____
49. Cuando me uno a un nuevo grupo, normalmente encajo pronto.
1. Verdadero 2. No estoy seguro 3. Falso _____
50. Me considero una persona socialmente muy atrevida y comunicativa.
1. Verdadero 2. No estoy seguro 3. Falso _____
51. Cuando estoy con un grupo, normalmente me siento, escucho y dejo que los demás lleven el peso de la conversación.
1. Verdadero 2. No estoy seguro 3. Falso _____
52. En ocasiones dudo si la gente con quien estoy hablando se interesa realmente por lo que digo.
1. Sí 0. No _____
53. Algunas personas parecen ignorarme o evitarme, aunque no sé por qué.
1. Verdadero 2. ? 3. Falso _____
54. A veces no puedo dormir porque tengo una idea que me da vueltas en la cabeza.
1. Verdadero 2. ? 3. Falso _____
55. Me pongo algo nervioso ante animales salvajes, incluso cuando están encerrados en fuertes jaulas.
1. Sí 2. ? 3. No _____
56. Cuando algo realmente me pone furioso, suelo calmarme muy pronto.
1. Sí 0. No _____
57. Me cuesta bastante hablar delante de un grupo numeroso de personas.
1. Verdadero 2. No estoy seguro 3. Falso _____
58. Algunas veces me "sacan de quicio" de un modo insoportable pequeñas cosas, aunque reconozca que son triviales.
1. Sí 0. No _____
59. En mi vida personal consigo casi siempre todos mis propósitos.
1. Verdadero 2. ? 3. Falso _____
60. Si cometo una falta social desagradable puedo olvidarla pronto.
1. Sí 2. ? 3. No _____

SEPARA ESTA HOJA DEL RESTO Y LEE LAS INSTRUCCIONES QUE SE DETALLAN SEGUIDAMENTE

A continuación queremos hacerte una pregunta referente a alguna de las cuestiones que acabas de responder. Se trata de los ítems de las páginas 2, 4, 6 y 8.

Habrás comprobado que junto a cada una de esas cuestiones aparece una línea. Si has respondido a alguna cuestión rodeando con un círculo la alternativa (2), indica en esa línea la RAZÓN que te ha llevado a hacerlo. Para ayudarte a realizar esta tarea te ofrecemos abajo una lista de posibles RAZONES. Si tu respuesta ante una cuestión ha sido la opción (2), escribe en la línea que aparece a la derecha de esa cuestión el número de la razón que refleja mejor por qué has contestado (2). Por ejemplo:

1. Me gusta presenciar una competición deportiva.

1. Verdadero

(2) Término medio

3. Falso

_____ 7 _____

Con esta respuesta (7), elegida de la lista que aparece abajo, estarías indicando que no te interesa la cuestión que se plantea y que por eso has elegido la opción (2).

Lee atentamente las razones de la lista, vuelve a la PÁGINA 2 donde comienzan las cuestiones y comprueba cuál ha sido tu respuesta a cada cuestión. Si has elegido la alternativa (2), escribe en la línea que aparece a la derecha del ítem el número de la razón que refleje mejor por qué has elegido dicha alternativa.

Realiza esta operación para TODAS LAS CUESTIONES DE LAS PÁGINAS 2, 4, 6 Y 8. Si no has respondido utilizando la alternativa 2, no escribas nada en la línea que aparece junto al ítem.

LISTA DE POSIBLES RAZONES

1. No entiendo bien la cuestión que se plantea
2. No quiero revelar cuestiones personales
3. Me falta información para elegir alguna de las otras dos opciones
4. Me siento en el punto medio que hay entre ambas opciones
5. Podría contestar cualquiera de las otras dos opciones porque ambas son aplicables a mí
6. Ninguna de las otras dos opciones me describe adecuadamente
7. No me interesa la cuestión, me siento indiferente
8. Otra razón (especifica cuál en la línea correspondiente)