

TEMA **1**

INTRODUCCIÓN A LA PSICOMETRÍA

María Isabel Barbero García

SUMARIO

1. Orientaciones didácticas
2. Aproximación al concepto de Psicometría
3. La Psicometría en el marco de la metodología de la Psicología como ciencia del comportamiento
4. La medición psicológica
5. Orígenes y desarrollo de la Psicometría
6. Los métodos de escalamiento
7. Desarrollo de los métodos de escalamiento de estímulos
 - 7.1. Métodos de escalamiento psicofísico
 - 7.2. Métodos de escalamiento psicológico
8. Orígenes y desarrollo de los tests
 - 8.1. La importancia de los trabajos de Galton
 - 8.2. Los primeros tests mentales
 - 8.3. Desarrollo de los primeros tests de inteligencia
 - 8.4. Los tests colectivos
 - 8.5. Las baterías de aptitud múltiple
 - 8.6. Los tests de personalidad
 - 8.7. La medición de los intereses y actitudes
 - 8.8. La institucionalización del uso de los tests
 - 8.9. Los tests referidos al criterio frente a los referidos a normas
 - 8.10. Los tests adaptativos informatizados (TAI s)
9. Desarrollo de la Teoría de los Tests
 - 9.1. Teoría Clásica de los Tests (TCT)
 - 9.2. Teoría de Respuesta al Ítem (TRI)
10. Ejercicios de autoevaluación
11. Soluciones a los ejercicios de autoevaluación
12. Bibliografía complementaria

1. ORIENTACIONES DIDÁCTICAS

En este primer tema se intenta dar una visión global de la Psicometría para que los alumnos se familiaricen con la disciplina cuyo estudio van a abordar; para que conozcan no sólo lo que es la Psicometría y cuál ha sido su evolución y desarrollo a lo largo de los años, sino para que comprendan el papel que desempeña en el proceso de investigación científica.

Para ello, después de hacer una revisión del concepto de Psicometría, y ofrecer una definición para mayor clarificación del mismo, se hace un análisis del área de conocimiento denominada *Metodología de las Ciencias del Comportamiento* en la que está incluida la Psicometría y que coincide con el nombre del Departamento. Este análisis va a permitir a los alumnos comprender la importancia que tienen algunas de las asignaturas que han de estudiar en el Grado de Psicología como son: *Introducción al Análisis de datos*; *Fundamentos de Investigación*; *Diseño y Análisis de datos* y, finalmente, *Psicometría*, puesto que les van a proporcionar las claves para poder llevar a cabo cualquier proceso de investigación científica.

Partiendo de que el concepto de medición es algo intrínseco a la Psicometría, y aceptando la necesidad y posibilidad de llevar a cabo mediciones en el marco de la Psicología como ciencia positiva que es, en este tema se hace un análisis de los dos caminos a través de los cuales se fue desarrollando la Psicometría: *Los Estudios de Psicofísica* y *los Estudios de las Diferencias Individuales*. Los primeros dieron lugar a los *Métodos de Escalamiento de estímulos* y los segundos al *Método de los Tests* para el escalamiento de los sujetos. En un principio estas dos vías de desarrollo siguieron caminos muy diferenciados pero, hoy día, esta separación está superada y suele mantenerse únicamente por motivos didácticos.

Teniendo en cuenta que la asignatura de Psicometría es una asignatura cuatrimestral dentro del plan de estudios de nuestra Universidad, ha sido necesario seleccionar los temas a incluir en este libro dada la amplitud de contenidos que abarca nuestra disciplina, y se ha creído conveniente centrarnos fundamentalmente en el estudio de la Teoría de los Tests. Por eso, no se hace una revisión

exhaustiva de los distintos métodos desarrollados para el escalamiento de los estímulos, sino un breve apunte, y sí una revisión más extensa del origen y desarrollo de los tests como instrumentos de medida y de las distintas teorías de los tests, para poder abordar en los temas que siguen el problema de su construcción, evaluación y aplicación.

Una vez estudiado este tema, los alumnos deberán tener muy claro el importante papel que juega la Psicometría en el marco de la Psicología científica. Deberán conocer los caminos a través de los que se fue desarrollando nuestra disciplina, sabiendo diferenciar entre los distintos métodos de escalamiento según que el objetivo perseguido sea el escalamiento de los estímulos, el de los sujetos o el de ambos (escalamiento de las respuestas).

Deberán aprender a conocer y valorar a figuras tan importantes como Fechner, Stevens y Thurstone por sus aportaciones para la elaboración de escalas psicofísicas y psicológicas y a figuras como Galton, Pearson, Cattell, Binet, Terman y tantos otros, por sus trabajos pioneros en el campo de la medida de las diferencias individuales, trabajos que dieron lugar al desarrollo del método de los tests y sentaron las bases para su evolución hasta alcanzar las cotas de desarrollo que tienen hoy en día, por ejemplo, los tests adaptativos informatizados.

Otro de nuestros objetivos es que los alumnos entiendan que los tests, como cualquier otro instrumento de medición, son imperfectos, y que las puntuaciones que obtienen los sujetos cuando se les aplican no representan con exactitud su verdadera puntuación en aquello que se está midiendo ya que están afectadas de errores de medida cuya cuantía es necesario estimar.

Deberán conocer las distintas teorías de los tests que se han ido desarrollando a lo largo de los años y que incluyen una serie de modelos matemáticos (o funciones). Estos modelos permiten establecer una relación entre las puntuaciones empíricas que obtienen los sujetos en los tests y su verdadera puntuación (o su nivel real) en la característica que se desea medir, y hacer estimaciones acerca de los errores presentes en el proceso de medición. Las diferentes funciones han dado lugar a diferentes teorías de los tests siendo las dos más importantes la Teoría Clásica de los Tests (TCT) y la Teoría de Respuesta al Ítem (TRI). Se puede hacer referencia también a la Teoría de la Generalizabilidad (TG) que, a pesar de que hoy en día no tiene la relevancia que se le auguraba en un principio, supuso un esfuerzo para tratar de dar solución a algunos de los problemas que quedaban sin resolver en la TCT.

2. APROXIMACIÓN AL CONCEPTO DE PSICOMETRÍA

Antes de comenzar el estudio de cualquier disciplina es necesario tener un conocimiento claro de qué es lo que se va a estudiar y porqué. Por lo tanto, dado que los contenidos se refieren a la Psicometría, el primer paso que hemos de dar es conceptuar el término y explicar lo que se en-

tiende por Psicometría, teniendo en cuenta, además, que no se trata de un concepto estático sino algo dinámico que irá evolucionando gracias a las aportaciones de las investigaciones realizadas en su campo, y se ampliará en la medida en que se vayan incorporando nuevos conocimientos.

La aproximación al concepto de la Psicometría, como al de cualquier otra disciplina, es una tarea difícil dada la variedad de facetas que presenta; sin embargo, si se quiere hacer un estudio riguroso del mismo puede abordarse a través de distintos caminos. Cada uno de estos caminos ofrecerá una información parcial; por eso, en la medida en que se utilice un mayor número de ellos se dispondrá de una información más completa.

Uno de los caminos a seguir para aproximarnos al concepto de Psicometría, y quizás el más inmediato, puede ser el análisis etimológico del término formado por las palabras griegas «*Psykhe*» y «*Metrum*», que literalmente significa «*Medida de la Psykhe*». Sin embargo, dado que la información que ofrece esta vía es bastante amplia y ambigua es necesario completarla; por eso se ha hecho un estudio de las definiciones dadas por algunos expertos (Cerdá, 1970; Cliff, 1979; García-Cueto, 1993; Macià, 1982; Martínez Arias, 1995; Melia, 1990; Morales, 1975; Muñiz, 1998; Musso, 1970; Nunnally, 1973; Rivas, 1976; Santisteban, 1990; Seoane, 1980 y Yela, 1968, entre otros) y, mediante un análisis de su contenido, se puede concluir que:

La Psicometría es una disciplina metodológica, dentro del área de la Psicología, cuya tarea fundamental es la medición o cuantificación de las variables psicológicas con todas las implicaciones que ello conlleva, tanto teóricas (posibilidades y criterios de medición) como prácticas (cómo y con qué se mide).

Analizando la definición anterior, se puede decir que la Psicometría deberá ocuparse en primer lugar de la *justificación y legitimación de la medición psicológica*, para lo cual deberá: a) desarrollar modelos formales que permitan representar los fenómenos que se quieren estudiar y posibiliten la transformación de los hechos en datos, b) validar los modelos desarrollados para determinar en qué medida representan la realidad que pretenden y c) establecer las condiciones que permitan llevar a cabo el proceso de medición. En segundo lugar, deberá también ocuparse de las *implicaciones prácticas y aplicadas que dicha medición conlleva*, proporcionando los métodos necesarios que indiquen, en cada caso concreto, cómo se debe llevar a cabo la cuantificación, y construyendo los instrumentos necesarios y adecuados para poder efectuarla. Esta vertiente aplicada de la Psicometría, referida a la construcción y evaluación de los instrumentos de medición, que es de la que nos vamos a ocupar fundamentalmente en este libro, no debe confundirse con el uso que se haga de los instrumentos una vez construidos. Un instrumento puede estar bien construido y, sin embargo, ser utilizado de forma incorrecta.

Si se quiere medir de alguna manera la extraversión, será necesario desarrollar el instrumento científico adecuado para llevar a cabo el proceso de medición, esa es la parte que le incumbe al

psicómetra; otra cosa muy diferente es el uso, bueno o malo, que se haga del instrumento una vez construido.

La importancia de la Psicometría, como disciplina a la que incumbe todo aquello relacionado con la medición de variables psicológicas, se justifica si se tiene en cuenta que en Psicología, como en las demás ciencias empíricas, el objetivo final es la *descripción, explicación y predicción* de los fenómenos de interés (en nuestro caso los fenómenos psicológicos) y se podrá cumplir mejor dicho objetivo mediante el proceso de medición. Por eso, aunque la Psicometría no tiene un campo de aplicación específico como sucede con otras disciplinas, su campo de aplicación abarca todos los campos de la Psicología: personalidad, procesos cognitivos, actitudes, etc., y juega un papel importantísimo ya que contribuye a fundamentar, elaborar y contrastar todas las teorías psicológicas. Es dentro de este marco donde se justifica la medición.

Ahora bien, las mediciones llevadas a cabo sin un contexto teórico o aplicación práctica que les sirva de base rara vez justifican el tiempo y el dinero que se invierte en ellas. Es necesario evitar caer en la tentación de considerar que la medición es la piedra de toque de la respetabilidad científica; como señala Miller (1982):

«...muchos psicólogos se han precipitado a buscar números antes de saber lo que esos números pueden significar... Todavía se pueden encontrar psicólogos que llevan a cabo mediciones de gran complicación y exageradamente precisas sólo para demostrar hasta qué punto el psicólogo puede llegar a ser científico. Hay gente que no admite que, si apenas merece la pena hacer una cosa, hacerla bien sigue sin merecer la pena» (pág. 115).

3. LA PSICOMETRÍA EN EL MARCO DE LA METODOLOGÍA DE LA PSICOLOGÍA COMO CIENCIA DEL COMPORTAMIENTO

En el apartado anterior se definió la Psicometría como una disciplina metodológica dentro del área de la Psicología; por eso, a la hora de adscribir las asignaturas del plan de estudio a un Departamento, la Psicometría lo fue al de *Metodología de las Ciencias del Comportamiento*.

Si se hace un análisis de esta denominación, nos encontramos con dos conceptos fundamentales: el concepto de *Metodología* y el de *Ciencias del Comportamiento*. En la medida en que puedan ser aclarados, se tendrá un mejor conocimiento del marco en el que se sitúa la Psicometría.

Partiendo de un análisis etimológico del término, *Metodología* significa *Tratado de los Métodos*, y teniendo en cuenta que dentro del marco de la ciencia el término *Método* hace referencia al camino que se debe seguir para poder conseguir el objetivo de la ciencia, se puede considerar que:

...la Metodología estudia las estrategias y procedimientos que, de una forma más o menos estructurada, se utilizan para la obtención de los conocimientos que constituyen una disciplina científica.

Por otra parte,

...las Ciencias del Comportamiento son aquellas que estudian la «conducta» mediante la utilización del método científico, con el fin de encontrar estructuras generales o leyes.

En este contexto el término «conducta» se utiliza en sentido amplio, y hace referencia a la actividad de un organismo ante una situación concreta que estará determinada biológica y socialmente.

Teniendo en cuenta que el método científico es el método común a todas las ciencias, que proporciona un marco general a cualquier proceso de investigación científica, pero que puede adaptarse a las peculiaridades de cada una de ellas en función de sus problemas específicos y de su objeto de estudio:

La Metodología de las Ciencias del Comportamiento estará referida al estudio del método general de la ciencia y de las estrategias o métodos específicos que deberán desarrollar cada una de ellas, en función de sus peculiaridades, para poder llevar a cabo su tarea.

La Psicología, como ciencia del comportamiento, tiene su propio objeto de estudio y sus propios problemas; por eso, deberá adaptar el método general de la ciencia, el método científico, al marco concreto de cada uno de los problemas, desarrollando las estrategias y técnicas complementarias que le permitan abordar su estudio y tratar de darles solución. A estas técnicas algunos autores las denominan *Técnicas metódicas* (Cruz Hernández, 1976; Fernández Ballesteros, 1980; Fernández-Trespalacios, 1979 y Moreno, 1983 entre otros).

Dentro del marco de la Metodología de las Ciencias del Comportamiento, y más concretamente de la Metodología de la Psicología como ciencia del comportamiento que es, encontramos una gran cantidad de contenidos relacionados entre sí por su carácter procedimental. Aunque estos contenidos por razones didácticas, y para que los alumnos aprendan a relacionar los contenidos entre sí, se hayan articulado en cuatro asignaturas dentro del Grado de Psicología, en realidad se pueden considerar tres grandes bloques cuyos orígenes van unidos a los de la Psicología científica: *Diseños de investigación, Análisis de datos y Psicometría*.

El bloque dedicado a los *Diseños de investigación* atiende, fundamentalmente, a aquella faceta de la investigación científica cuya tarea fundamental es la operativización de las variables de la hipótesis y la elaboración de un plan de trabajo, o procedimiento para la recogida de los datos, que sea coherente con la hipótesis; puesto que, de acuerdo con Arnau (1990) el concepto de diseño

de investigación *está esencialmente vinculado a la elección y especificación del procedimiento para la obtención de los datos relevantes a la hipótesis* (pág. 13). Para llevar a cabo esta tarea, es necesario analizar, entre otros, los siguientes aspectos: la naturaleza de las variables, sus posibilidades de manipulación, la elección de aquellas que sean de interés para la investigación objeto de estudio, la detección de variables extrañas y formas de control, los criterios de selección y asignación de unidades de observación, la estructuración interna del procedimiento con delimitación de situaciones, tareas, etc. (Sarriá, 1991). En la medida en que se resuelvan todos estos aspectos se reducirá la incertidumbre que conlleva la elección de un diseño de investigación.

El bloque correspondiente al *Análisis de datos* será el encargado de proporcionar las técnicas necesarias para llevar a cabo el tratamiento estadístico de los mismos, tratamiento que puede ir desde la simple descripción o representación gráfica a procedimientos más complejos de ajuste de modelos o contrastes de hipótesis.

Por último, la *Psicometría*, comparte con el resto de las disciplinas psicológicas tanto el objeto de estudio: la conducta humana, como el método: el método científico; entonces, lo que realmente la caracteriza es la peculiar adaptación que hace de éste al objeto de estudio de la Psicología. Al incluir todo lo referente a la medición, la Psicometría proporciona las reglas que van a permitir llevar a cabo el proceso de operativización de las variables que se quieren medir. Una vez obtenidas las medidas mediante la asignación de números, los modelos psicométricos permitirán un análisis del error que les afecta (fiabilidad de las medidas) y, finalmente, los estudios de validación permitirán hacer inferencias acerca de las relaciones entre los datos empíricos obtenidos (medidas) y el constructo o variable psicológica que se quiere medir.

Como se desprende de todo lo anterior, a pesar de que estos tres bloques tienen una entidad propia, cada uno tiene una enorme influencia sobre los demás, y juntos interactúan en el proceso global e integrado que es la investigación científica.

Nota: Creo que esta pequeña introducción permitirá a los alumnos conocer la relación que existe entre las asignaturas del Departamento de Metodología de las Ciencias del Comportamiento y su importancia dentro de los estudios de Grado de Psicología.

4. LA MEDICIÓN PSICOLÓGICA

Dado que en las asignaturas de *Introducción al Análisis de Datos y Fundamentos de Investigación* ya se abordó el tema de la medición en Psicología no vamos a extendernos nosotros en su exposición, pero sí queremos resaltar la importancia del mismo dado que se trata de un problema inherente al desarrollo de la Psicometría y de la Psicología científica. Hasta el momento en que

se acepta la posibilidad de medir lo psicológico no se consideraba que la Psicología fuera una ciencia.

De acuerdo con Coombs, Dawes y Tversky (1981), consideramos que uno de los papeles fundamentales asignados a la Ciencia es la descripción, explicación y predicción de los fenómenos observables por medio de unas cuantas leyes generales que expresen las relaciones entre las propiedades de los objetos investigados. En las Ciencias más avanzadas las leyes expresan relaciones cuantitativas, lo cual indica que las propiedades de los objetos se pueden representar por medio de números mediante un proceso de *medición*. La Psicología como Ciencia tendrá su base científica en la medición, que le permitirá contrastar empíricamente las hipótesis planteadas.

Para Nunnally (1970), la medición se reduce a algo muy sencillo, *consiste en un conjunto de normas para asignar números a los objetos de modo tal que estos números representen cantidades de atributos* (pág. 23), *entendiendo por atributos las características de los objetos y no los objetos mismos*. Cuando decimos que queremos medir una mesa, lo que realmente queremos hacer es medir alguna característica de la mesa, como por ejemplo su longitud o su altura, no la mesa como objeto. Esto implica que la medición conlleva un proceso de abstracción. Hoy día la medición se toma, en general, como la asignación de números a entidades, acontecimientos o sucesos, con el fin de representar sus propiedades y sus relaciones.

Ya en la antigüedad, cuando se quería medir algún atributo físico de los objetos, como podía ser su peso, su longitud, etc., se trataban de desarrollar los instrumentos pertinentes y nadie dudaba acerca de la posibilidad de llevar a cabo tales mediciones. Sin embargo, la polémica surgió cuando en lugar de querer medir atributos físicos se intentaron medir atributos psicológicos, puesto que existían serias dudas acerca de la posibilidad de llevar a cabo tales mediciones. Una de las formas de paliar esas dudas era mostrar que las variables psicológicas se podían cuantificar y que los procedimientos que se utilizaban para ello permitían establecer relaciones cuantitativas entre las variables.

No obstante, a nadie se le escapa la dificultad que entrañaba la medición de características psicológicas dada la singularidad de las mismas y, por lo tanto, las dificultades que hubo que ir superando hasta que se consiguió que se aceptara la necesidad y posibilidad de medir este tipo de variables. La dificultad principal derivaba, fundamentalmente, de que a diferencia de los atributos físicos de los sujetos, como el peso y la estatura, que pueden ser medidos directamente con los instrumentos pertinentes, la mayoría de los atributos psicológicos como por ejemplo la inteligencia, el autoritarismo y la introversión, son conceptos abstractos, denominados también *constructos teóricos (variables latentes)*, cuya medición no puede llevarse a cabo de forma directa sino que debe inferirse a través de la medición de una serie de conductas representativas de dicho constructo. En este sentido, Zeller y Carmines (1980) plantearon una nueva concepción de la medición; consideraron que se trataba de *un proceso mediante el cual se enlazan conceptos abstractos (los constructos inobservables directamente), con indicadores empíricos observables directamente (las conductas)*.

Si por ejemplo se quiere medir la *inteligencia*, lo primero que se nos ocurre preguntar es ¿qué es la inteligencia? Entonces nos daremos cuenta de que es algo abstracto, inobservable y que, por lo tanto, no se puede medir directamente. Sin embargo, estamos acostumbrados a oír decir de las personas que son poco o muy inteligentes lo cual quiere decir que, de alguna manera, se puede evaluar eso que nosotros llamamos inteligencia. Una forma de hacerlo puede ser analizar cómo se comportarían las personas, a las que consideramos inteligentes, ante determinadas situaciones y, posteriormente, crear el instrumento adecuado para medir esas conductas. De esta manera, la variable psicológica *inteligencia*, que es un constructo teórico inobservable de forma directa, se puede manifestar a través de una serie de conductas que ya sí son observables directamente y, por lo tanto, pueden ser medidas mediante el instrumento adecuado. A partir de los resultados obtenidos en el proceso de medición, se podrán hacer inferencias acerca del grado en que cada uno de los sujetos evaluados posee el constructo de interés, en nuestro caso, del grado de inteligencia que poseen.

Este tipo de medición se suele denominar *medición por indicadores* entendiendo que, dado que las variables psicológicas no se pueden medir de forma directa, es necesario seleccionar una serie de indicadores que sí pueden ser medidos directamente, y que se supone están en estrecha relación con el constructo o variable psicológica que se quiere medir.

Hoy día se han desarrollado distintos procedimientos para la medida de las sensaciones, aptitudes, actitudes, etc., pero hasta hace poco no tenían la fundamentación lógica necesaria para su justificación. Como señala Muñiz (1992), los únicos métodos que había para evaluar la calidad métrica de las medidas obtenidas eran una colección de técnicas estadísticas encuadradas bajo las denominaciones de *fiabilidad* y *validez*; de ahí que no era de extrañar que las mediciones llevadas a cabo en el campo de la Psicología fueran consideradas, tanto cualitativamente como cuantitativamente, de orden inferior a las realizadas en el campo de la física. Sin embargo, los desarrollos recientes han demostrado que, aunque las medidas obtenidas al medir variables psicológicas puedan ser menos precisas que las realizadas en el campo de la Física el estatus teórico de las mediciones puede justificarse al mismo nivel (pág. 232).

Además del problema anterior, a la hora de medir variables psicológicas nos encontramos con otro problema, el de las unidades de medida que se van a utilizar al interpretar las puntuaciones obtenidas por los sujetos. Si nosotros medimos la longitud de una mesa el resultado lo podemos expresar en centímetros, si lo que evaluamos es el peso de cualquier objeto el resultado se expresará en gramos o kilogramos, pero si lo que medimos es la inteligencia de un niño o su capacidad para las matemáticas, ¿qué unidades de medida podemos utilizar para dar un significado a los resultados obtenidos?

En Psicología, hay dos formas fundamentales de abordar el problema, una *referida a normas* y otra *referida al criterio*. La forma más habitual de proceder es la primera, la referida a normas, que consiste en comparar los resultados obtenidos por ese niño con los obtenidos por un grupo de

niños que forman el llamado *grupo normativo* y que pertenecen a su misma población; es decir, que pertenecen a su misma clase, son de su misma edad, etc. En otras ocasiones la interpretación se hace en relación a un criterio previamente establecido; los resultados obtenidos se comparan con ese *criterio* (un punto crítico) y la superación o no del mismo es lo que va a dar significado a las puntuaciones obtenidas.

Las dificultades que entraña la medición psicológica se comprenden mejor si, como señala Muñoz (1998), se tiene en cuenta que la conducta humana se desarrolla en una banda acotada por una base neurobiológica y un entorno sociocultural y surge, por lo tanto, de la interacción entre nuestra constitución biológica y la estimulación ambiental.

Nota: El lector interesado en profundizar en el problema de la medición en Psicología puede consultar, entre otros, el libro de José Luis Meliá (1990) *Introducción a la medición* y el de Joel Michell (1999), *Measurement in Psychology: a critical history of a methodological concept*.

Sea cual sea el campo de aplicación de la medición psicológica (procesos básicos, personalidad, procesos cognitivos, actitudes, valores, etc.) hay una serie de objetivos comunes fundamentales: en primer lugar estimar los errores aleatorios que conlleva toda medición (fiabilidad de las medidas) y garantizar que la misma no es algo inútil sino que sirve para explicar y predecir los fenómenos de interés (validez de las medidas). Todos estos aspectos se irán analizando a lo largo de los temas siguientes, después de haber hablado acerca de los orígenes y vías de desarrollo de la Psicometría, o lo que es lo mismo de los orígenes y vías de desarrollo de la medición psicológica.

5. ORÍGENES Y VÍAS DE DESARROLLO DE LA PSICOMETRÍA

Partiendo de que es necesaria y posible la medición en Psicología, y teniendo en cuenta que la Psicometría es la disciplina que entre sus contenidos incluye todo lo relativo a la medición, vamos a ir analizando brevemente los hechos fundamentales que contribuyeron a su desarrollo y convirtieron a la Psicología en una ciencia positiva independiente de la Filosofía.

De acuerdo con Yela (1968), podemos decir que existieron dos motivos fundamentales que posibilitaron la introducción de la medición en Psicología. Uno de ellos, la tendencia a formular los problemas científicos en términos matemáticos; el otro motivo, el enfrentamiento de la Psicología, hacia la mitad del siglo XIX con dos problemas fundamentales: el primero, el estudio cuantitativo de las relaciones entre las características físicas de los estímulos y las sensaciones que suscitan en los sujetos en función de las cuales se asignan valores numéricos a los estímulos: *el problema psi-*

cofísico; el segundo, el problema de la *cuantificación de las diferencias individuales* mediante la asignación de números a los sujetos en función del grado en que manifiesten un atributo o conducta.

Así pues, se puede situar el origen de la Psicometría hacia la mitad del siglo XIX y, a partir de ese momento, se va a desarrollar, fundamentalmente, a través de estas dos vías:

- a) los estudios de Psicofísica que dieron lugar al desarrollo de modelos que permitieron asignar valores numéricos a los estímulos y, por lo tanto, que permitieron *el escalamiento de estímulos*.
- b) los estudios acerca de las diferencias individuales que dieron lugar al desarrollo de los Tests y de las distintas Teorías de los Tests (modelos), que posibilitaron la asignación de valores numéricos a los sujetos y, por lo tanto, *el escalamiento de los sujetos*.

Estos dos puntos de arranque de la Psicometría (los estudios de psicofísica y la cuantificación de las diferencias individuales) dieron lugar al desarrollo de las dos corrientes que mayor incidencia han tenido en la investigación psicológica: la experimentalista y la correlacional.

En la Psicometría clásica se reservaba el término *Escalamiento* para designar el proceso de construcción de escalas para la cuantificación de estímulos; mientras que todo lo relativo a la cuantificación de los sujetos, que se estudiaba en la *Teoría de los Tests*, no se consideraba parte de dicho proceso. En la actualidad, la distinción entre *Escalamiento* y *Teoría de los Tests*, aunque se mantiene por cuestiones didácticas, está superada gracias al desarrollo de nuevas teorías de la medición y de numerosas técnicas estadísticas comunes a ambas vertientes de la Psicometría.

6. LOS MÉTODOS DE ESCALAMIENTO

Podemos considerar el escalamiento como el campo de la Psicometría cuyo objetivo fundamental es la construcción de escalas de medida; es decir, la construcción de instrumentos que permitan llevar a cabo mediciones para representar las propiedades de los objetos (estímulos, sujetos o respuestas) por medio de números, de acuerdo a unas normas o reglas.

Dentro del marco de la Psicometría, hay dos supuestos básicos en todos los métodos de escalamiento:

- La existencia de un continuo latente o subyacente, a lo largo del cual varían los objetos psicológicos que se van a escalar (estímulos, sujetos, o respuestas), que no puede ser observado de forma directa.
- Que los objetos psicológicos (estímulos, sujetos o respuestas) pueden situarse de forma ordenada a lo largo de ese continuo.

Por ejemplo, si el atributo o característica que se quiere medir es la inteligencia, supondremos que ese atributo se puede representar a lo largo de un continuo (una recta en términos geométricos), sobre el cual se podrán situar los sujetos de forma ordenada en función del grado de inteligencia que manifiesten (siguiendo con el símil de la representación geométrica, los sujetos se podrían representar a lo largo de la recta por medio de puntos).

Hemos estado considerando como objetos psicológicos a los estímulos, a los sujetos y a las respuestas. Ahora bien ¿qué diferencias nos vamos a encontrar en función de que los objetos psicológicos a escalar sean estímulos, sujetos o respuestas?

Como señalaron Ghiselli, Campbell y Zedeck (1981, pág. 392) *los estímulos son las cosas que los investigadores presentan normalmente a un sujeto con el propósito de elicitar una respuesta*. Partiendo de esta definición, el término estímulo tiene un sentido muy amplio ya que pueden ser considerados como tales no sólo un conjunto de objetos físicos, sino una lista de adjetivos que hagan referencia a algún rasgo de personalidad, una serie de frases que se refieran a alguna variable de actitud, un grupo de personas a las que se vaya a calificar, una serie de problemas de Psicometría, etc.

En cualquier caso, el propósito del escalamiento de estímulos es determinar las características que los sujetos perciben en ellos y, por lo tanto, la respuesta del sujeto (o sujetos) ante la presentación de los estímulos es una respuesta subjetiva que nos va a permitir diferenciarlos y escalarlos; es decir, asignar un valor numérico a cada uno de los estímulos. Cuando se utilizan varios sujetos para el escalamiento de estímulos, el valor que se asigna a cada uno de ellos suele venir determinado por algún índice de tendencia central, la media o la mediana, obtenido a partir de los valores asignados por cada uno de los sujetos a cada estímulo. En este caso serían los estímulos los que se situarían a lo largo del continuo en función de sus valores escalares (los valores asignados) y los sujetos que han emitido juicios acerca de esos estímulos, los que habrían actuado como instrumentos de medida.

Cuando el objeto a escalar son los sujetos se utiliza una muestra, generalmente extraída de forma aleatoria de una población, y todos los sujetos que la componen responden al mismo conjunto de estímulos (los elementos de un test, por ejemplo); posteriormente, las respuestas emitidas por los sujetos ante la presentación de cada uno de los estímulos serán combinadas de alguna manera para proporcionar una puntuación numérica para cada sujeto de la muestra. Las diferencias encontradas en las puntuaciones obtenidas reflejarán las diferencias entre los sujetos respecto al atributo o característica que se está midiendo. Se asume que los estímulos presentados son interpretados de la misma forma por todos los sujetos de la muestra y, por lo tanto, las variaciones encontradas son debidas a las diferencias entre los sujetos. Serán los sujetos los que se situarán a lo largo del continuo que representa la característica que se está midiendo, y los estímulos los que habrán actuado como instrumento de medida.

Hay veces que lo que interesa es situar sobre el continuo, a lo largo del cual varía el atributo que se está estudiando, tanto a los estímulos como a los sujetos. En este caso, las variaciones en-

contradas en las respuestas de los sujetos ante la presentación de los estímulos se atribuyen, no sólo a las diferencias que hay entre los estímulos respecto al grado de atributo que llevan implícito, sino también a las diferencias que hay entre los sujetos en cuanto a su actitud personal frente a cada uno de los estímulos. En estos métodos la posición de cada sujeto respecto del atributo que se está analizando, su actitud, aptitud, sentimientos, etc., es un factor que está determinando su respuesta. Esta tercera aproximación a los métodos de escalamiento fue denominada por Torgerson (1958) *aproximación centrada en la respuesta*.

Vemos, por lo tanto, que los métodos de escalamiento hacen referencia tanto al escalamiento de estímulos como al de sujetos, o al de ambos a la vez.

7. DESARROLLO DE LOS MÉTODOS DE ESCALAMIENTO DE ESTÍMULOS

Dado que en sus orígenes el escalamiento estuvo asociado al escalamiento de estímulos vamos a mantenerlo aquí de esa forma por razones didácticas, tal y como ya se ha comentado, y abordaremos el estudio de todo lo referente al escalamiento de sujetos dentro del marco de la teoría de los tests.

El origen de los métodos de escalamiento de estímulos tuvo lugar cuando la Psicología se enfrentó con el problema de cuantificar, de alguna manera, las relaciones existentes entre las características físicas de los estímulos y las sensaciones que dichos estímulos suscitan en los sujetos; es decir, cuando la Psicología se enfrenta con el problema psicofísico.

7.1. Métodos de escalamiento psicofísico

A principios del siglo XIX el filósofo y pedagogo alemán Herbart acuñó el concepto de *umbral mínimo* para designar a la mínima intensidad que tiene que tener un estímulo para que se pueda percibir; vemos ya, por lo tanto, un primer intento de relacionar la intensidad de los estímulos con las sensaciones que producen. A partir del concepto de umbral mínimo Weber (1795-1878) comenzó a desarrollar una serie de procedimientos experimentales, *los métodos psicofísicos*, que iban a permitir el cálculo de los umbrales y desarrolló su famosa *Ley de Weber*, que establece que:

... el incremento de magnitud que debe experimentar un estímulo (ΔE) para que el sujeto perciba que se ha producido un cambio, es una proporción constante de su magnitud inicial (E):

$$K = \frac{\Delta E}{E}$$

[1.1]

A la constante K se la conoce como constante de Weber; el (ΔE) es el incremento mínimo que ha de experimentar la magnitud de un estímulo (magnitud física) respecto a la magnitud inicial (E) para que el sujeto perciba un cambio mínimo en la sensación (magnitud psicológica), a este cambio mínimo de sensación le denominó Weber *diferencia apenas perceptible* (dap).

Aunque a Weber se deben los primeros intentos de establecer una ley general para formular la noción de umbral, la figura más representativa fue Fechner (1801-1887), que desarrolló una serie de métodos denominados *métodos psicofísicos indirectos* que permitían elaborar unas escalas denominadas *escalas psicofísicas* (Baird y Noma, 1978; Barbero, 1993/1999, 2007; Blanco, 1996; Fechner, 1966; Muñiz, 1991, entre otros).

Supongamos que se quieren escalar una serie de estímulos, que varían en cuanto a su peso, respecto al grado de pesadez que producen en los sujetos. En este caso, cuando a los sujetos se les presenten los distintos estímulos a escalar, deberán emitir un juicio acerca del grado de pesadez que han percibido al sopesar cada uno de ellos. El peso es una característica física de los estímulos que varía a lo largo de un continuo físico y hay instrumentos adecuados para su medición; por el contrario, la pesadez es una característica psicológica o subjetiva que varía a lo largo de un continuo psicológico, y los sujetos actúan como un instrumento de medida asignando valores a cada uno de los estímulos en función del grado de pesadez que les hayan provocado al presentárselos. Estos valores son psicológicos o subjetivos.

Por lo tanto, en el escalamiento psicofísico se tienen dos continuos, un continuo físico a lo largo del cual varían los estímulos y uno psicológico a lo largo del cual variarán las sensaciones que dichos estímulos producen en los sujetos. Para construir la escala psicofísica, será necesario ver qué relación funcional se puede establecer entre los dos continuos.

Según Fechner, la función que mejor representa la relación entre los dos continuos es una *función logarítmica* cuya fórmula es:

$$S = C \ln E + A$$

[1.2]

Siendo:

S = valor en la escala de sensación.

E = valor del estímulo.

C y A = la pendiente y la ordenada en el origen de la función logarítmica.

Para poder medir la relación entre la magnitud de los estímulos y las sensaciones que producen Fechner introdujo los conceptos de *umbral absoluto* y *umbral diferencial*. Para Fechner, el *umbral absoluto* sería la magnitud física del estímulo que se requiere para que se produzca una sensación y el *umbral diferencial* sería el incremento mínimo, en la magnitud física del estímulo, que se requiere para que el sujeto perciba un cambio de sensación.

La ley de Fechner establece que cuando la magnitud física del estímulo está en el umbral absoluto la sensación es nula, en ese punto se establece el origen en la escala de sensación (será el 0 de la escala de sensación), y que si se aumenta la estimulación en proporción geométrica las sensaciones aumentarán aritméticamente. Es decir, que cada vez se necesitará un mayor incremento en la estimulación física para que se perciba un cambio en la sensación.

Los principales supuestos en los que se basa la ley de Fechner son los siguientes:

1. Asume la ley de Weber:

$$K = \frac{\Delta E}{E}$$

Es decir, que el incremento de magnitud que debe experimentar un estímulo (ΔE) para que el sujeto perciba que se ha producido un cambio es una proporción constante de su magnitud inicial (E).

2. Asume que todas las diferencias apenas perceptibles (d.a.p) son psicológicamente iguales. Esto implica que siempre que se produzca un aumento en la magnitud del estímulo igual a un umbral diferencial (ΔE), sea cual sea el valor de esa magnitud, la sensación aumentará siempre en la misma cuantía. Esto puede expresarse así: $\Delta S = I$, donde I es una constante.
3. Establece que el origen de la escala de sensación, es decir el punto cero de la misma, corresponde al umbral absoluto, es decir al valor del estímulo correspondiente al valor absoluto en la escala física.

EJEMPLO:

Supongamos que se quiere medir la capacidad auditiva de una persona; para ello, se le presentan una serie de estímulos de distinta intensidad comenzando por uno cuya intensidad de sonido esté por debajo de la capacidad auditiva de la persona. Poco a poco se va aumentando la intensidad del sonido hasta que la persona empiece a detectar el estímulo. En ese momento ha habido un cambio, la persona ha pasado de no tener ninguna sensación a percibir el sonido. El valor del estímulo que ha provocado ese cambio corresponde al *umbral absoluto* de la persona y marca el origen de la escala de sensación (el punto cero). Si a partir del valor correspondiente al umbral absoluto se va aumentando muy lentamente la magnitud del estímulo (en nuestro ejemplo la intensidad del sonido), llega un momento en que la persona detecta que ha habido un cambio y se pro-

duce en ella un cambio de sensación. En ese momento se ha encontrado su primer *umbral diferencial*, que correspondería al incremento mínimo que tiene que experimentar la intensidad del estímulo para que la persona perciba un cambio de sensación. Al cambio de sensación se le denomina *diferencia apenas perceptible (dap)*, por lo que se dirá que se ha producido una (dap) a partir del umbral absoluto.

Siguiendo con el mismo razonamiento se irían calculando los diferentes umbrales diferenciales que permitirán obtener dos series de valores, una correspondiente a los distintos valores del estímulo (escala física), y la otra serie la escala de sensación (escala psicológica) cuyos valores se obtendrán a base de ir sumando las *dap* que se han ido produciendo a partir del origen. Una vez obtenidas las dos series de valores sería necesario comprobar la relación que existe entre ellas y si esta relación es logarítmica como postula Fechner.

Si representamos gráficamente sobre unos ejes de coordenadas los pares de valores obtenidos se observará la relación funcional que existe entre ellos y que según Fechner es logarítmica.

Supongamos el caso hipotético de que el umbral absoluto de una persona para el peso son 10 gramos; es decir, que hasta que no se le presente al sujeto un estímulo cuyo peso sean 10 gramos el sujeto no percibe ninguna sensación, supongamos también que para que el sujeto note que hay un cambio en el peso del estímulo se necesita aumentar 2 gramos el peso inicial, lo cual supone que el umbral diferencial de ese sujeto y para ese estímulo inicial es de dos gramos, bajo estos supuestos la constante de Weber sería $K = 0,2$ dado que:

$$K = \frac{2}{10} = 0,2$$

Si se cumplieran los supuestos de la ley de Fechner obtendríamos los siguientes valores para el estímulo (continuo físico) y la sensación (continuo psicológico):

<i>E</i>	<i>S (dap)</i>
10	0
12	1
14,4	2
17,28	3

¿Cómo se han obtenido estos valores?

Al umbral absoluto le corresponde el cero en la escala de sensación. Si tal y como hemos comentado anteriormente, para que la persona note un cambio de sensación es necesario aumentar el peso del estímulo en 2 gramos, la constante de Weber será 0,2 y el umbral diferencial 2 gramos.

Por lo tanto, cuando el valor del estímulo pasa de 10 a 12 gramos se ha producido en el sujeto 1 (dap), por eso en la escala psicológica aparece el 1.

Sabiendo que la constante de Weber es $K = 0,2$ y que la ley de Fechner asume la ley de Weber, podemos ir averiguando los distintos valores tanto de la escala física como de la psicológica. El siguiente valor será: $0,2 = \frac{\Delta E}{12} \rightarrow \Delta E = 0,2 \times 12 = 2,4$

Cuando el valor del estímulo es de 12 gramos y la constante de Weber 0,2 se necesita un aumento de 2,4 gramos para que el sujeto perciba un cambio en la sensación. Por lo tanto al valor del estímulo igual a 14,4 (12 + 2,4) en la escala física le corresponde el valor de 2 en la escala de sensación puesto que desde el umbral absoluto se han producido 2 dap.

De esta manera se irían obteniendo los valores de ambas escalas suponiendo que los datos se ajustaran a la función de Fechner.

Si se quiere calcular el valor de un estímulo cualquiera, el n-esimo, se debería aplicar la siguiente fórmula:

$$E_n = E_{n-1} + (E_{n-1})K = E_{n-1}(1 + K) \quad [1.3]$$

En el ejemplo que hemos puesto no sería necesario averiguar el ajuste de los datos a la función logarítmica ya que, de antemano, sabemos que el ajuste es perfecto puesto que los hemos ido obteniendo asumiendo los supuestos de la ley de Fechner; ahora bien, el problema que hay que abordar es el de comprobar si, en situaciones reales, se verifica la ley.

La forma de comprobarlo es sencilla, en primer lugar hay que seleccionar los valores del estímulo y, posteriormente, por medio de una serie de métodos llamados métodos psicofísicos se irán obteniendo los datos y se irá elaborando la escala de sensación. Una vez obtenidas las dos escalas lo único que hace falta comprobar es si los datos obtenidos se ajustan o no a la función logarítmica de Fechner.

Ahora bien, el problema fundamental está en la obtención experimental de los umbrales absoluto y diferencial que permitirán medir la capacidad de detección y de discriminación de los sujetos. Para ello, Fechner desarrolló una serie de procedimientos entre los que destacaremos: *el método de los límites, el método de ajuste y el método de los estímulos constantes*. El procedimiento general que se sigue para llevar a cabo el escalamiento (procedimiento que variará en función del método utilizado) consiste en presentar a una muestra de sujetos el conjunto de estímulos a escalar y, en función de las respuestas emitidas por aquellos, a lo largo de los distintos experimentos, se asignará un valor numérico a cada uno de los estímulos.

En el *método de los límites*, también llamado de *los cambios mínimos*, es el experimentador el que va modificando la intensidad del estímulo. Cuando se quiere calcular el umbral absoluto co-

menzará, o bien por un estímulo de baja intensidad e irá aumentándola hasta que el sujeto que los va a evaluar comience a detectarlo, o bien por uno de alta intensidad e irá reduciéndola poco a poco hasta que el sujeto deje de percibirlo. En el caso de querer calcular los umbrales diferenciales el experimentador seleccionará un valor del estímulo como estándar y, junto a él, irá presentando al sujeto un estímulo de comparación cuya intensidad irá manipulando hasta que el sujeto considere que la intensidad del estímulo de comparación es igual que la del estímulo estándar.

En el *método de ajuste* también llamado de *error promedio*, la diferencia con respecto al método anterior radica en que, en lugar de ser el investigador el que manipula la intensidad de los estímulos, es el propio sujeto el que la va modificando, aumentándola o disminuyéndola, hasta encontrar el valor de sus umbrales.

En el *método de los estímulos constantes* se asume que cuando un estímulo se presenta a un mismo sujeto en repetidas ocasiones no siempre es percibido y que aún cuando lo perciba no siempre le produce la misma sensación. Partiendo de este supuesto, para averiguar el umbral absoluto, cada estímulo se presenta a los sujetos un número elevado de veces en orden aleatorio y, en cada ocasión, cada sujeto ha de decir si lo ha percibido o no. El umbral absoluto es la magnitud del estímulo que ha sido percibido por los sujetos el 50% de las veces que se ha presentado. Para el cálculo del umbral diferencial, al igual que en los métodos anteriores, se fija un valor del estímulo como estándar y, a continuación, se van presentando una serie de estímulos de comparación cuyo valor estará situado simétricamente en torno al del estándar. Se repetirán varias veces los ensayos y, en cada uno de ellos, el sujeto deberá decir si el estímulo de comparación es mayor o menor que el estándar. Este método es el más utilizado para el cálculo de los umbrales absoluto y diferencial (ver Barbero 1993/1999; 2007 para el cálculo empírico de los umbrales).

La ley de Fechner, a pesar de su indudable importancia, pronto recibe varias críticas debidas, fundamentalmente, a que los estímulos de muy alta o muy baja intensidad no se ajustan bien a ella y, por otra parte, tampoco lo hacen todos los sistemas sensoriales. No obstante, es indudable que introdujo una nueva forma de medición en Psicología, de ahí que se considere que con Fechner comienza la etapa de la Psicología cuantitativa que se ha aplicado a toda clase de problemas psicológicos. Sus trabajos, junto con los que Wundt llevó a cabo en su laboratorio de Psicología fundado 19 años después, marcaron el comienzo de la *Psicología experimental*.

Las controversias que surgen a partir de la Psicofísica desarrollada por Fechner influyeron en otra de las figuras más relevantes en el campo de la Psicofísica: Stevens (1906-1972), quién hace una reformulación de la Psicofísica fechneriana desarrollando los métodos de escalamiento que esta nueva reformulación requería, y una nueva función conocida con el nombre de *Función potencial de Stevens*:

$$R = q \cdot E^n$$

[1.4]

Siendo:

R = respuesta de los sujetos.

E = valor del estímulo.

q = constante que depende de las unidades de medida.

n = exponente de la función que depende del atributo sensorial.

Los trabajos llevados a cabo por Stevens contribuyeron, en gran medida, al resurgimiento de la investigación en el campo de la Psicofísica pues ponían a prueba y comparaban los dos tipos de funciones, la logarítmica y la potencial.

A los métodos desarrollados por Stevens se les conoce también como *métodos directos de escalamiento*, ya que en lugar de originar una escala de sensación (S), como ocurría con los métodos de Fechner, dan lugar a una escala de respuesta (R) cuyos valores son las estimaciones subjetivas que, de forma directa, hacen los sujetos de los estímulos presentados, y cuyas propiedades van a estar determinadas por las instrucciones dadas a los sujetos, a los que se considera capaces de hacer estimaciones a nivel de intervalo y razón.

Stevens trata de medir por un lado el estímulo (E) y por otro la respuesta que de forma directa emiten los sujetos para, posteriormente, encontrar una función que los relacione. Según Stevens esta función es una función potencial. A diferencia de Fechner, no basa sus mediciones en ninguna suposición acerca de las ($d\alpha p$), que Fechner utiliza como unidad de medida de su escala de sensación, ni en el concepto de umbral.

Entre los métodos de escalamiento que utiliza Stevens, merecen citarse: *métodos de emparejamiento de magnitudes* (por modalidad cruzada, estimación de magnitudes y producción de magnitudes), *métodos de emparejamiento de razones* (por modalidad cruzada, estimación de razones y producción de razones), *métodos de emparejamiento de intervalos* (por modalidad cruzada, estimación de intervalos y producción de intervalos) y *las escalas de categorías o de clasificación* (*Rating scales*).

En los **métodos de emparejamiento de magnitudes por:**

Modalidad cruzada:

Se utilizan dos continuos físicos diferentes y la tarea del sujeto consistirá en emparejar uno con el otro. El experimentador selecciona una serie de estímulos, que varían a lo largo de un continuo determinado cuya magnitud se puede medir mediante los instrumentos adecuados, por ejemplo la intensidad de un sonido, y la tarea de cada sujeto consistirá en ajustar, a cada uno de los estímulos presentados, otro estímulo que varía a lo largo de otro continuo, por ejemplo la separación entre dos rectas verticales y paralelas. Para obtener los valores, tanto de la escala correspondiente al estímulo (escala E) como los correspondientes a la respuesta del sujeto (escala R), se dispone de los instrumentos de medida adecuados.

Estimación de magnitudes:

Se presenta a cada sujeto un estímulo y se le advierte que se fije en él puesto que va a servir de estímulo estándar a partir del cuál habrá de estimar los valores correspondientes al resto de los estímulos que se le van a presentar. En este método, puede ser el experimentador el que asigne un módulo al estímulo estándar y el sujeto irá asignando valores al resto de los estímulos que se le vayan presentando, tomando como referencia el valor asignado al estímulo estándar. Supongamos que se está haciendo un estudio sobre la percepción de los sujetos acerca de la longitud de una serie de líneas y al estímulo que se presenta como estándar se le asigna un módulo de 8 cm. Si el estímulo presentado a continuación le parece al sujeto que es la mitad de largo que el presentado como estándar deberá asignarle un valor de 4 cm si, por el contrario, le parece el doble de largo deberá asignarle un valor de 16 cm.

Producción de magnitudes:

En este método la tarea del sujeto es inversa respecto al método anterior. El experimentador va presentando al sujeto una serie de números, de uno en uno y de forma aleatoria, y la tarea del sujeto consiste en modificar la magnitud de los estímulos en base a los números presentados por el experimentador, de manera que cada número lleve emparejado un estímulo cuya magnitud será, a juicio del sujeto, la representada por el número.

En los métodos de emparejamiento de razones por:*Modalidad cruzada:*

Se presentan al sujeto dos estímulos que guardan una determinada proporción entre ellos y se le pide que ajuste otros dos, pertenecientes a otro continuo diferente, de manera que guarden entre sí la misma proporción que guardaban los dos primeros. Supongamos que el experimentador le presenta al sujeto dos rectas de la misma longitud y le pide que produzca dos ruidos cuya intensidad mantenga la misma proporción que la que mantenía la longitud de las rectas entre sí; es decir, que sean de la misma intensidad.

Estimación de razones:

Se le presentan al sujeto todos los pares de estímulos, y su tarea consiste en hacer estimaciones de las razones que hay entre las magnitudes de cada par y asignar un número a cada par que represente esa razón. Se trata de asignar razones numéricas a las razones entre las magnitudes de los estímulos. Si mantenemos el ejemplo de la longitud entre dos rectas y en uno de los pares que se le presentan al sujeto la primera línea le parece que es la mitad de larga que la segunda, deberá asignar a ese par el valor numérico de $\frac{1}{2}$, ya que esta razón numérica es la que, a juicio del sujeto, representa la razón entre la longitud de las líneas que forman el par presentado.

Producción de razones:

Se presenta al sujeto un estímulo estándar y junto a él una proporción numérica. La tarea del sujeto consiste en producir otro estímulo que guarde con el estándar una proporción igual a la presentada. Supongamos que al sujeto se le presenta una línea recta de una longitud determinada y se le pide que produzca un estímulo cuya longitud sea la mitad que la del anterior, o bien el doble, etc.

En los métodos de emparejamiento de intervalos por:*Modalidad cruzada:*

Dados una serie de estímulos que varían a lo largo de dos continuos divididos en intervalos el sujeto habrá de emparejar los intervalos existentes entre los estímulos de un continuo con los intervalos existentes entre los estímulos del otro continuo.

Estimación de intervalos:

Se le presentan al sujeto una serie de estímulos diferentes entre sí y se le pide que, mediante números, haga una estimación de las diferencias estimulares.

Producción de intervalos:

Se le presentan al sujeto dos estímulos y su tarea consiste en encontrar un estímulo intermedio entre los dos presentados (bisección), una serie de estímulos que dividan el intervalo entre los dos estímulos presentados en más de dos intervalos iguales (equisección) o en una serie de intervalos distintos (multisección).

El método de **escalas de categorías o clasificación** es uno de los más utilizados tanto en Psicología como en Sociología. Consiste en asumir que el continuo a lo largo del que se han de situar los estímulos está dividido en una serie de categorías ordenadas cuyos límites serán fijos salvo por error aleatorio. La tarea a realizar será estimar los valores escalares de los límites de las categorías para, una vez hecho esto, asignar los estímulos a cada una de ellas y averiguar sus valores escalares.

A diferencia de los métodos de Fechner que permitían medir la capacidad de detección (mediante umbrales absolutos) y de discriminación (mediante los umbrales diferenciales) de las personas, los métodos desarrollados por Stevens están centrados en medir la capacidad de los sujetos para hacer estimaciones subjetivas acerca de la magnitud de los estímulos y comprobar hasta qué punto los juicios emitidos (estimaciones subjetivas) se ajustan a los datos reales (magnitud real de los estímulos). De esta manera se puede formalizar algo que estamos haciendo constantemente en nuestra vida diaria y que, sin embargo, no siempre somos conscientes de ello. Cuántas veces hemos dicho u oído frases como las siguientes:

Mi casa mide unos 300 metros cuadrados.

Hace aproximadamente dos horas.....

Pues bien, lo que se pretende es comprobar hasta qué punto esas respuestas emitidas por los sujetos se ajustan a la realidad de los datos.

7.2. Métodos de escalamiento psicológico

Los métodos de escalamiento se desarrollaron, en principio, para su utilización en el campo de la percepción pero utilizando características de los estímulos que variaban a lo largo de alguna dimensión física y que, por lo tanto, podían ser medidas con los instrumentos adecuados; se trataba de establecer una relación entre las características físicas de los estímulos y las sensaciones que producían. Ahora bien, ¿qué ocurriría si la característica que se quisiera escalar no variara a lo largo de ningún continuo físico? Si, por ejemplo, se quisiera medir el grado de realismo de una serie de cuadros, las preferencias políticas de una muestra de sujetos, la agresividad, la actitud de los españoles ante la inmigración, etc., difícilmente se podría llevar a cabo el proceso de medición utilizando ninguna escala física, puesto que estas características no varían a lo largo de ningún continuo físico sino a lo largo de un continuo psicológico.

Al surgir este problema comienzan a desarrollarse a finales del siglo XIX, y sobre todo durante el siglo XX, una serie de métodos de escalamiento que aunque basados en las ideas de Fechner tienen unas características propias, como son la no necesidad de recurrir a medidas de tipo físico. A este tipo de métodos de escalamiento se les agrupa bajo la denominación genérica de *métodos de escalamiento psicológico* y a las escalas resultantes *escalas psicológicas o subjetivas*. La figura que más contribuyó al desarrollo de este tipo de escalas fue Thurstone (1887-1955).

Thurstone nació en Chicago de padres suecos, y después de obtener el título de ingeniero civil en la universidad de Cornell se dedicó a la electrotecnia. Antes de graduarse había patentado un modelo de proyector de películas que atrajo la atención de Edison quien le ofreció un puesto en su laboratorio de East Orange (Nueva Jersey), pero Thurstone pronto abandona el laboratorio para dedicarse a enseñar geometría y dibujo. En 1914 se interesa por el estudio experimental del aprendizaje y se matricula en Psicología en la Universidad de Chicago siendo ayudante de Bingham en el Departamento de Psicología Aplicada del Instituto Carnegie de Tecnología, del que llegó a ser Director después de doctorarse.

A finales de la década de 1920, L. L. Thurstone estaba en la universidad de Chicago como profesor de Psicofísica, pero convencido de que las sensaciones que medía con tanta precisión no merecían su tiempo y esfuerzo, y aburrido de todo aquello que enseñaba, decidió probar fortuna y medir, aplicando los mismos métodos, algo que para él tuviera importancia.

En lugar de presentar a los sujetos dos objetos y preguntarles, por ejemplo, ¿cuál de estos dos objetos es más largo?, se les podrían presentar dos cuadros y preguntarles ¿cuál de estos dos cua-

dros tiene más realismo?, o bien, tal y como hizo, presentarles una serie de delitos y pedirles que los diferenciaron en función de su gravedad. Si este tipo de problemas se pudiera tratar mediante algún tipo de lógica psicofísica, se abriría la posibilidad de una descripción objetiva de mayor significación psicológica que el umbral sensorial (Miller, 1982).

Thurstone trató de elaborar un modelo a partir del cual pudiera elaborar una escala sobre un continuo psicológico y situar en ella los estímulos (también psicológicos) sin necesidad de recurrir a ninguna operación en un continuo físico. El modelo que desarrolló está basado, por una parte, en la variabilidad perceptual de los sujetos, e incluso en la de un mismo sujeto cuando se le presentan los mismos estímulos en distintas ocasiones y, por otra, en la limitación que tienen los sujetos para percibir las diferencias de magnitud entre dos estímulos cuando estas son muy pequeñas. En la medida en que la diferencia entre la magnitud de los estímulos sea mayor, será más fácil que los sujetos puedan diferenciarlos y ordenarlos respecto a la característica o atributo que se esté evaluando; mientras que, en la medida en que los estímulos sean más parecidos, los sujetos encontrarán mayores dificultades para realizar su tarea. Thurstone (1927) publicó varios trabajos que trataban diversos problemas acerca de la medición subjetiva o psicológica, y desarrolló un modelo matemático, relacionado con la Psicofísica clásica de Fechner, cuyas ecuaciones se conocen con el nombre de *Ley del Juicio Comparativo*. Posteriormente desarrollaría otro modelo cuyas ecuaciones se conocen como *Ley del Juicio Categórico*.

Estos dos modelos llevan asociados una serie de métodos experimentales para la obtención empírica de los datos; la Ley del Juicio Comparativo utiliza el *Método de las Comparaciones Binarias*. La Ley del Juicio Categórico: el *Método de los Intervalos Sucesivos*, el *Método de los Intervalos Aparentemente Iguales* y el *Método de Ordenación de Rangos* fundamentalmente.

Aunque Thurstone atribuyó a otros la fundación de la *Psychological Society* y de su revista *Psychometrika*, se hallaba muy unido al grupo de personas que en 1936 crearon la sociedad y la revista para fomentar el desarrollo de la Psicología como ciencia racional cuantitativa.

A partir de los trabajos de Thurstone se fueron desarrollando nuevas formas de escalamiento psicológico, así:

Guttman en los años 40 del siglo xx, desarrolló un nuevo modelo para el escalamiento conjunto de sujetos y estímulos (escalamiento de respuestas). Para la obtención de la escala utilizó un método conocido como *método del escalograma*. La escala resultante se denomina *escala de entrelazamiento* puesto que tanto los sujetos como los estímulos se sitúan a lo largo del mismo continuo psicológico de forma entrelazada.

Coombs (1950), desarrolló una teoría conocida como *Teoría del despliegue* y propuso un modelo que, al igual que el de Guttman, permite escalar sujetos y estímulos conjuntamente.

Todos estos modelos de escalamiento y sus métodos asociados se desarrollaron, en principio, para la construcción de escalas unidimensionales; es decir, de escalas que permitieran ordenar un

conjunto de estímulos, o estímulos y sujetos en los dos últimos casos, respecto a un único atributo o característica mediante la asignación de un único valor escalar que represente la posición del objeto escalado.

En los años 60 se desarrollaron los métodos de *escalamiento multidimensional*, aunque, en realidad, los orígenes de estos métodos haya que situarlos en 1938 en los tratados de Young y Hosenholder y los de Richardson. A diferencia de los métodos de escalamiento unidimensionales, estos métodos permiten la ordenación de los objetos a escalar atendiendo simultáneamente a más de un atributo o característica y, por lo tanto, asumen la existencia de más de una dimensión subyacente al conjunto de observaciones. En este caso, en lugar de asignar un único valor en la escala a cada uno de los objetos, se les asignará un valor en cada una de las dimensiones analizadas.

Hay tres obras clásicas sobre escalamiento, tanto psicofísico como psicológico. La obra de Guilford *Psychometric Methods* publicada en 1954; la de Torgerson *Theory and Methods of Scaling* de 1958 y el libro de Edwards *Techniques of Attitude Scale Construction* de 1957.

En castellano se puede consultar el libro de Barbero (1993/1999; 2007) y el de Blanco (1996).

8. ORÍGENES Y DESARROLLO DE LOS TESTS

Como se comentó al inicio del tema, el segundo problema que motivó la introducción del proceso de medición en Psicología fue el intento de apreciar de forma sistemática las diferencias individuales; es decir, el intento de escalar a los sujetos. Este intento llevó al desarrollo del *Método de los Tests* y de las distintas *Teorías de los Tests*.

Podríamos remontarnos bastantes años antes de Cristo y ya se apreciaría el interés existente por analizar de alguna manera las diferencias individuales. En China se utilizaban tests para seleccionar a las personas que ocuparían puestos en el gobierno. Autores clásicos como Platón e Hipócrates propusieron también algunas formas para conseguir analizar las diferencias individuales y, de un modo más concreto, en el siglo XVI lo hace el español Huarte de San Juan.

En 1796 Kinneybrook, uno de los investigadores que trabajaba en el Observatorio Astronómico de Greenwich, fue expulsado porque discrepó del resto de sus compañeros en la estimación que hizo del tiempo que tardaba una estrella en cruzar un determinado espacio, y se consideró que se había equivocado en los cálculos. Hubo que esperar hasta el año 1822 a que los astrónomos comprendieran que las personas tienen diferentes tiempos de reacción y que, por lo tanto, a la hora de interpretar los datos observados era necesario tener esto en cuenta (Freeman, 1926), y hubo que esperar hasta la mitad del siglo XIX para que se desarrollaran en Europa y en Estados Unidos procedimientos formalizados para la evaluación de tales diferencias: *los tests*.

Antes de continuar con la reseña histórica del origen y desarrollo de los tests creemos necesario aclarar el significado psicométrico del término. Son muchas las definiciones que han ido apareciendo a lo largo de los años, pero todas ellas hacen referencia a que:

Un test es un instrumento de medición diseñado especialmente para estudiar de un modo objetivo y sistemático el nivel de los sujetos respecto a algún atributo, característica o dominio de conocimientos y, a partir de las puntuaciones que obtengan los sujetos en el test, poder analizar las diferencias existentes entre ellos.

Hay tres factores que se pueden considerar decisivos en el desarrollo de los tests: la apertura del laboratorio antropométrico de Galton en Londres, el desarrollo de la correlación de Pearson y la interpretación que Spearman hace de ella, considerando que la correlación entre dos variables indica que ambas tienen un factor común.

8.1. La importancia de los trabajos de Galton

Francis Galton (1822-1911), era nieto de Erasmus Darwin y medio primo de Charles Darwin, por lo que es fácil comprender la influencia que tuvo la teoría de la evolución en sus trabajos.

Las ideas evolucionistas plasmadas en la obra de Charles Darwin titulada *The Origin of Species by Means of Natural Selection* y publicada en 1859, tuvieron una enorme acogida por parte de Francis Galton pero preocuparon enormemente a Wundt ya que el evolucionismo defendía una filosofía completamente diferente a la suya. Mientras que Wundt trataba de encontrar las leyes generales que dirigen o gobiernan la *mente* humana, los evolucionistas trataban de clasificar y catalogar las diferentes formas de *mentes* que podían existir. Esto fue el desencadenante que propició la separación entre la Psicología Experimental y la Psicología diferencial.

Galton todavía creía que la medición de las características mentales estaba estrechamente relacionada con la de las características físicas, y pensaba que para poder estudiar las dimensiones de la mente del hombre debía obtener la misma clase de datos antropológicos que estaba reuniendo al tratar de estudiar su anatomía. Comprendió que para obtener una descripción completa del hombre, las medidas antropométricas de estatura, peso, color de la piel, etc., deberían completarse con medidas psicométricas de los sentidos, la memoria, etc. Pero, como señala Miller (1982) hasta el día de su muerte fue reacio a admitir que *el tamaño del cráneo de un hombre no tenía valor como medida de su inteligencia*.

En 1884 montó su laboratorio antropométrico en la Exposición Internacional de la Salud celebrada en el Museo de South Kensington en Londres. Allí, por el precio de tres peniques, se tomaban a los visitantes medidas en una variedad de pruebas físicas y sensoriales pues, entre otras cosas, consideraba que la inteligencia podía medirse a través de la actividad sensomotora y, una vez

recogidos los datos, comenzó elaborando distribuciones de frecuencias y comprobando que su forma era aproximadamente la misma para las variables psicológicas que para las anatómicas, la distribución normal. Ahora bien, a cada persona se le habían tomado datos de distintas variables y, ante esa cantidad de datos, pronto se plantea el problema de cómo encontrar relaciones entre medidas diferentes. ¿Cuál era la relación entre ellas? Por ejemplo la gente de estatura alta tendía a pesar más que la de estatura baja. ¿Cómo podía medirse esta tendencia? Este problema, que podía tener cierta lógica, se complicaba cuando se trataba de hacer estudios sobre la herencia. ¿Cómo se podían relacionar las mediciones hechas a los padres con las de los hijos? Para solucionar este problema Galton recurrió a la correlación como medida de asociación. Así, Galton fue el primero en aplicar el concepto estadístico de distribución normal, de media, de mediana, varianza y correlación a datos psicológicos. Esta tarea fue continuada por Karl Pearson (Miller, 1982).

Para poder llevar a cabo sus mediciones, Galton construyó y utilizó una serie de instrumentos que pueden ser considerados los primeros *Tests psicométricos*, pero éstos se caracterizaban por un fuerte carácter antropométrico, sensorial y motor. No obstante, a partir de Galton empieza el gran desarrollo del método de los tests.

8.2. Los primeros tests mentales

James McKeen Cattell en 1888, estuvo en la Universidad de Cambridge y allí se puso en contacto con Galton, estableciéndose entre ellos un nexo de unión dado su común interés por investigar las diferencias individuales; no obstante, este trabajo común duró poco tiempo ya que Cattell volvió a Estados Unidos como profesor de Psicología en la Universidad de Pennsylvania. En 1891 se trasladó a Columbia donde fundó un laboratorio de Psicología en el que trabajó durante 26 años, hasta que fue dimitido por sus posiciones pacifistas cuando Estados Unidos entró en la Primera Guerra Mundial en 1917.

La asociación entre Cattell y Wundt en Alemania y de Cattell y Galton en Inglaterra fue, por una parte, el lazo de unión entre los laboratorios psicológicos alemanes e ingleses y, por otra, un hecho que contribuyó al intercambio de ideas entre los investigadores europeos y americanos (French y Hale, 1990).

Cattell utilizó por primera vez el término *Test Mental* en su artículo *Mental Test and Measurement* publicado en 1890 en la revista *Mind*. Pero los tests a los que hace referencia Cattell, al igual que los de Galton, tenían un carácter sensorial y motor fundamentalmente, ya que incluían medidas de energía muscular, velocidad de movimientos, sensibilidad al dolor, etc. El análisis de los datos puso de manifiesto que la correlación entre este tipo de pruebas y el nivel intelectual de los sujetos era nula (Wissler, 1901). Para Cattell los tests constituían *un sistema uniforme que permite comparar y combinar, en lugares y momentos diferentes, la medida de las funciones mentales* (Cattell, 1890;

pág. 374), y compartía con Galton la creencia de que se podía medir el funcionamiento intelectual de las personas mediante tests de discriminación sensorial y midiendo el tiempo de reacción.

Aunque durante las dos últimas décadas del siglo XIX hay una enorme expansión de los estudios acerca de las diferencias individuales en Inglaterra y en Estados Unidos, los tests seguían siendo, fundamentalmente, de tipo sensorial y motor. Fue necesario llegar a finales de siglo para aceptar que estos tests no medían la inteligencia. Las puntuaciones obtenidas por los sujetos en estos tests no guardaban relación con las medidas de rendimiento escolar, lo que venía a refutar la hipótesis de la relación entre la capacidad sensorial y la inteligencia.

8.3. Desarrollo de los primeros tests de inteligencia

Alfred Binet (1857-1911), fue el primero en darse cuenta de que las sensaciones no jugaban un papel demasiado importante en la Psicología diferencial y que era necesario centrarse en el estudio de los procesos mentales superiores. Otro de los aciertos de Binet fue el darse cuenta también de la importancia que tenía la edad de los sujetos como variable interviniente. En 1889 fundó con Beaunis el primer laboratorio de Psicología de la Sorbona del que llegó a ser director en 1894.

Binet trabajó con Simon y realizaron una serie de investigaciones cuyo resultado fue la publicación en 1905 de lo que puede ser considerada la primera escala de inteligencia. A diferencia de los tests de Galton, este test ya no se ocupaba de las funciones motoras o sensoriales sino de la capacidad de comprensión y razonamiento de los niños. Esta escala, conocida como la *Escala de Binet-Simon* constaba de 30 elementos de dificultad creciente, y aunque se incluían algunos elementos de tipo sensorial la mayoría eran verbales. Esta escala fue revisada en 1908, traducida a numerosos idiomas y adaptada varias veces. La adaptación más conocida del test de Binet fue la que realizó en 1916 el psicólogo americano Terman en la Universidad de Stanford, de ahí el nombre de *Test Stanford-Binet de Terman*. Este test fue traducido y adaptado en España por Germain y Rodrigo (1928).

Aunque la finalidad inicial de la escala era detectar a los sujetos que presentaran algún retraso intelectual, posteriormente se utilizó para el estudio de las diferencias en otros niveles. Para poder interpretar las puntuaciones obtenidas, se desarrolló el concepto de *Edad Mental* que equivale a la edad cronológica de los niños intelectualmente normales cuya media en la escala es igual a la puntuación obtenida por el niño examinado. Es decir, se aplica la escala a una muestra representativa de niños intelectualmente normales divididos en distintos niveles de edad y se calcula la puntuación media de los niños en cada uno de los niveles. Esa sería la norma establecida con la cual comparar los resultados obtenidos. Así, supongamos que se aplica la escala a un niño de 12 años y obtiene una puntuación que coincide con la puntuación media obtenida por los niños de 9 años; entonces, diremos que la edad cronológica del niño examinado es de 12 años pero su edad mental es de 9 años.

Terman (1877-1956), consideraba que la inteligencia se podía definir como la habilidad para llevar a cabo razonamientos abstractos y, a pesar de que sigue la táctica de Binet para la elaboración de sus tests, utiliza como medida de la inteligencia el concepto de *Cociente Intelectual* (CI) acuñado por Stern en 1912. El cociente intelectual (no coeficiente intelectual como se escucha muchas veces) equivale al cociente entre la edad mental y la edad cronológica, generalmente multiplicado por cien para evitar los decimales. Como señala Santisteban (1990) al definir el cociente intelectual se establece la primera escala de puntuaciones referida a las aptitudes mentales. En 1937 junto a Merrill llevaron a cabo otra revisión del test de Binet y obtuvieron dos formas paralelas.

8.4. Los tests colectivos

Todos los tests desarrollados hasta el momento eran de aplicación individual, lo que representaba un gran inconveniente por la cantidad de tiempo que requería su aplicación; pero la entrada de EE.UU. en la Primera Guerra Mundial en 1917, y la necesidad de seleccionar y clasificar el contingente de personas disponibles en función de sus capacidades, hacía imposible la aplicación de este tipo de pruebas. Entonces, se nombró una comisión especial de la Asociación Norteamericana de Psicología, dirigida por R.M. Yerkes, con el encargo de investigar nuevos procedimientos que posibilitaran la administración de tests en el ejército. Para ello, se pusieron en contacto con *Arthur Otis* que llevaba tiempo trabajando en un test colectivo de inteligencia y, a partir de ese material, se diseñaron los famosos tests conocidos como *Tests Alpha* y *Beta* que fueron publicados en 1918. El primero, el Test Alpha, diseñado para la población general y el segundo, el Test Beta, para los reclutas analfabetos o aquellos que no dominaban el inglés. Ambos tests siguen usándose hoy día tras numerosas revisiones. Este fue el comienzo de los tests colectivos (Otis, 1939).

Una vez finalizada la guerra el uso de los tests se extendió a la industria y al resto de las instituciones. En 1922 Catell fundó la primera empresa dedicada a la publicación masiva de tests, y a partir de ese año empezaron a utilizarse normalmente como instrumento de selección en la administración americana. Su sistema de valoración era sencillo; en general, se asumía como puntuación de los sujetos el número de aciertos y para su interpretación se recurría, bien a normas cronológicas (edad mental o cociente intelectual), bien a normas estadísticas (centiles o puntuaciones típicas).

En 1938 apareció el *Test de Weschler-Bellevue* para la medición de la inteligencia en los adultos y en 1949 apareció una versión para niños conocida, de forma abreviada, como *WISC*. En 1955 apareció una revisión de la escala anterior denominada *Weschler Adult Intelligence Scale (WAIS)* que, a su vez, fue revisada en 1981, y en 1967 apareció el *Weschler Preschool and Primary Scale of Intelligence*. La aparición de estas escalas supuso un gran avance en el desarrollo de los tests, entre otras cosas, porque se había constatado la falta de efectividad que tenía el test de Stanford-Binet aplicado a adultos.

A pesar del gran desarrollo del método de los tests, quedaban una serie de preguntas sin respuesta: ¿qué es lo que miden realmente los tests?, ¿existen realmente los rasgos que dicen medir los tests? En un intento de dar respuesta a estos interrogantes se desarrollaron el coeficiente de correlación de Pearson y una serie de técnicas correlacionales conocidas bajo el nombre genérico de *Análisis Factorial*. Los primeros esbozos de estas técnicas se deben a Spearman y hay que resaltar la enorme importancia que tuvieron en el desarrollo del método de los tests, ya que sin estas técnicas los tests mentales hubieran tardado más en perder su carácter básicamente sensorial.

El objetivo común de las técnicas reunidas bajo el término *Análisis Factorial* es representar un conjunto de variables observables (escalas, ítems, etc.) mediante un número más pequeño de variables no observables sino latentes llamadas factores de las que las primeras son indicadores.

La influencia del Análisis Factorial respecto a los tests de inteligencia fue doble. Por una parte dio una fundamentación teórica a su utilización, mostrando que en la mayor parte de las funciones cognitivas interviene un factor general «g» común a todas ellas. Por otra, los resultados del análisis factorial subrayan la importancia de estudiar otras aptitudes más específicas en el campo de la inteligencia. Surgen así los primeros tests destinados a la medida de las aptitudes y del rendimiento.

8.5. Las baterías de aptitud múltiple

Los tests de inteligencia utilizados hasta el momento proporcionaban información acerca del nivel intelectual global de los sujetos y eso no bastaba, era preciso trazar su perfil mental en función de las puntuaciones específicas obtenidas en diversos factores cognoscitivos. Así, a partir de los años treinta, y sobre todo a partir de los cuarenta del siglo XX, cuando ya la técnica del análisis factorial empezaba a dar sus frutos, comenzaron a desarrollarse las baterías de tests (*Baterías de Aptitud Múltiple*), cuya finalidad era procurarnos una medida de la posición de cada sujeto en un cierto número de rasgos. En lugar de una puntuación total, o CI, se obtiene una puntuación por separado para cada rasgo, dando, como señala Muñiz (1992), una mayor importancia a un factor general de inteligencia que articularía jerárquicamente otros factores de grupo (escuela inglesa), o reclamando un plano de igualdad para factores múltiples (escuela americana).

Entre las primeras baterías merece destacar la *Batería de Aptitudes Mentales Primarias* de Thurstone (Thurstone, 1938; Thurstone y Thurstone, 1941) que incluía pruebas para la medida de lo que entonces se consideraban los componentes fundamentales del comportamiento inteligente: comprensión verbal, fluidez verbal, aptitud numérica, aptitud espacial, memoria, rapidez perceptiva y razonamiento general. Hoy día existen numerosas baterías (PMA, DAT, etc.) de uso habitual.

8.6. Los tests de personalidad

Aunque la medida de las diferencias individuales había estado centrada fundamentalmente en el campo de las facultades mentales; se habían hecho algunos intentos de extender el uso de los tests a otros campos de la Psicología: la personalidad, los intereses, las actitudes, etc. Se puede considerar a Kraepelin como un precursor de los *Tests de Personalidad*, pues ya en 1894 utilizó el *Test de Asociación Libre* con pacientes normales para estudiar los efectos psicológicos de la fatiga, el hambre y las drogas, y llegó a la conclusión de que todos esos agentes aumentan la frecuencia relativa de las asociaciones superficiales (Anastasi, 1980).

El prototipo de cuestionario de personalidad con carácter psicométrico, es la *Hoja de Datos Personales* de Woodworth (1917), utilizada en la Primera Guerra Mundial y cuya finalidad era la detección de aquellos sujetos que por padecer trastornos neuróticos graves no eran aptos para el servicio militar. Posteriormente se prepararon distintas formas de este cuestionario e incluso se preparó una forma para niños.

De los primeros tests objetivos de personalidad hay tres que han sobrevivido en la actualidad: una serie desarrollada por Guilford y sus colaboradores (Guilford y Zimmerman, 1949); una serie similar desarrollada por Cattell y sus colaboradores (Cattell, Eber y Tatsuoka, 1970) y el *Inventario Multifásico de Personalidad de Minnesota* (MMPI) de Hathaway y McKinley (1942). En general los tests de Guilford y de Cattell se usan en «poblaciones normales» mientras que el MMPI se usa en «poblaciones clínicas». Además, los tests de Guilford y de Cattell están basados en el análisis factorial y están orientados hacia los rasgos mientras que el MMPI, en su forma estándar, no utilizaba escalas derivadas del análisis factorial y está orientado hacia una clasificación psiquiátrica.

Durante la primera mitad del siglo xx se fueron perfilando otro tipo de tests, los *Tests Proyectivos*; así, en 1921, el psiquiatra suizo Rorschach diseñó el primer test que pretendía dar una visión global y comprensiva de la personalidad, el *Test de Psicodiagnóstico*, se trata del famoso test proyectivo también conocido como *el Test de las manchas de tinta*. A partir de este test se desarrollaron otros muchos que pronto se vieron asociados con la Psicología clínica y, en especial, con el Psicoanálisis. En 1926 aparece el test de *Dibujo de la Figura Humana* de Machover, en 1938 el *Test de Apercepción Temática* (TAT), etc.

No obstante, como señalan Goldstein y Hersen (1984), hay varias razones que han contribuido a que disminuya la utilización de estas técnicas: a) el aumento de la sofisticación científica ha creado una atmósfera de escepticismo en torno a estos instrumentos, b) el desarrollo de procedimientos alternativos, por ejemplo el MMPI y otros tests objetivos, convenció a algunos psicólogos clínicos de que a partir de estos procedimientos se podía obtener la misma información y de una manera menos costosa y c) las técnicas proyectivas, en general, estaban asociadas en cierta medida con la teoría psicoanalítica. A pesar de estas razones científicas, prácticas y filosóficas el test de Rorschach sigue mostrando su utilidad.

8.7. La medición de intereses y actitudes

Entre los instrumentos clásicos para la medida de los intereses merecen citarse el *Cuestionario de Intereses Vocacionales de Strong (SVIB)* desarrollado por E.K. Strong (1927) y la *Escala de Preferencias de Kuder* introducida por Kuder en 1934. La más reciente revisión de esta escala es la efectuada en 1985.

En cuanto a la medida de las actitudes, creencias y opiniones, su desarrollo efectivo tuvo lugar a partir de 1928, fecha en la que Thurstone desarrolló sus dos modelos, la *Ley del Juicio Comparativo* y, sobre todo, la *Ley del Juicio Categórico*, junto con los procedimientos experimentales para la recogida de datos que, basados en los métodos psicofísicos desarrollados por Fechner iban a permitir el escalamiento de los ítems (estímulos). La técnica utilizada por Thurstone para la medida de las actitudes (Thurstone y Chave, 1929), tal y como se verá más adelante, permite la ordenación de los ítems (estímulos) a lo largo de una escala continua en función de los valores escalares que les hayan asignado el grupo de sujetos a los que se les han presentado para su evaluación (prueba de jueces); una vez escalados los ítems se podrá utilizar la escala para averiguar el grado de actitud de los sujetos en la variable medida; es decir, para escalar a los sujetos.

Algunos años más tarde Likert (1932) desarrolló una nueva técnica que vino a paliar alguno de los inconvenientes de la de Thurstone, entre otros la prueba de jueces. Las escalas elaboradas siguiendo la *Técnica de Likert*, están basadas en los mismos principios que la mayoría de los tests de aptitudes. Algunos investigadores, entre los que podemos citar a Edwards y Kenney (1946), encontraron que este tipo de escalas era más fácil de elaborar que las de Thurstone y diagnosticaban mejor.

Con el fin de medir el significado connotativo, también llamado significado afectivo o subjetivo, que determinados estímulos tienen para los sujetos, Osgood (1952) revisó todas las teorías del significado y así pudo encontrar el marco teórico que le permitió desarrollar una escala de clasificación para medirlo: *El Diferencial Semántico*.

Nota: Dada la importancia de las contribuciones de Thurstone no vamos a extendernos aquí en la explicación de sus modelos y los métodos que llevan asociados. En el Tema 3 se hará una exposición algo más detallada de los mismos, junto a otras técnicas como las de Likert, Guttman y Osgood, por su importancia para la medida de las actitudes.

8.8. Institucionalización del uso de los tests

Durante los años cuarenta y cincuenta del siglo xx se produjo un aumento considerable en el uso de los tests pero también un cierto abuso, lo que condujo a numerosas críticas sociales e in-

cluso a la prohibición de su uso en determinados estados. Yela (1977) cuenta cómo en Estados Unidos se aplicaron durante 1946 unos sesenta millones de tests a cerca de veinte millones de personas. A las críticas sociales provocadas por los excesos en el uso de los tests se unieron ciertas críticas aparecidas en la literatura científica acerca de la adecuación de los mismos para ciertos usos y colectivos, dado que muchos de ellos presentaban determinados sesgos (Jensen, 1980; Levine y Rubin, 1979; Lumsden, 1976; Pawlik, 1976). No obstante, como señala Santisteban (1990) las limitaciones de los tests no justificaron del todo esta fuerte corriente crítica, puesto que la teoría y la práctica demostraron que:

... las deficiencias que a priori parecen ser propias de los instrumentos que tratan de medir capacidades humanas complejas a través de simples manifestaciones observables, son generalmente conocidas y controlables y que, por lo tanto, basta considerarlas para hacer un uso correcto del test. Asimismo, habrá que tenerlas en cuenta al hacer la interpretación de las puntuaciones, puesto que es obvio, en cualquier proceso de medición, el que la interpretación de la medida se ajuste también a las características del instrumento (pág.3).

De acuerdo con Meliá (1990) la cuestión puede consistir en no confundir la teoría psicométrica con los tests derivados de ella, ni los tests con las teorías psicológicas ni con los usos inadecuados que se hagan de los mismos.

El uso masivo de los tests fue extendiendo su campo de aplicación a distintos sectores tanto públicos como privados, en la selección de puestos en el gobierno y en las escuelas, para ayudar a niños con problemas escolares, en orientación, clasificación, etc., lo que ocasionó la aparición de instituciones tanto públicas como privadas dedicadas a ello. El aumento de estudiantes que querían acceder a la escuela secundaria hizo necesario un plan de selección, así en 1899 se formó el *College Entrance Examination Board (CEEB)* quienes desarrollaron una serie de tests para realizar dicha selección. Los primeros exámenes del CEEB estaban basados en contenidos curriculares, pero hacia el año 1928-29 se desarrolló el *Scholastic Aptitude Test (SAT)* que intentaba medir la *capacidad para el aprendizaje* más que lo que ya habían aprendido los estudiantes.

En 1947, el CEEB, el *American Council on Education* y otras instituciones se unieron para fundar el *Educational Testing Service (ETS)* con el fin de potenciar la investigación en el campo educativo. El ETS jugó un importante papel en el desarrollo histórico de la Psicometría como soporte y núcleo de formación de una gran parte de las principales figuras extranjeras que hoy día existen en este campo y de algunas españolas que han ido allí y siguen yendo a formarse.

En 1951 fueron nombrados el comité de la *American Psychological Association (APA)*, el de la *American Educational Research Association (AERA)*, y el *National Council on Measurements Used in Education (NCMUE* más tarde *NCME)* para preparar una serie de recomendaciones técnicas para el uso de los tests, tal y como se verá más adelante. Inicialmente, intentaron preparar cuatro

manuales independientes sobre aptitudes, rendimiento, intereses y personalidad basados en los trabajos de los diferentes comités, pero debido a la similitud de las recomendaciones se decidió que la APA publicara un volumen sobre aptitudes, intereses y personalidad y que la AERA y el NCME publicaran uno sobre rendimiento a través del *National Education Association*, (APA, comité sobre tests psicológicos, 1955). Publicación que constituye un punto de referencia común.

En la publicación llevada a cabo en 1999, la *American Psychological Association (APA)*, la *American Educational Research Association (AERA)*, y el *National Council on Measurements in Education (NCME)*, publicaron los *Standards for Educational and Psychological Testing*, con el fin de proporcionar a los profesionales de la Psicometría criterios para la evaluación de los tests. La última edición es la de 2014.

En España, y con el fin de proporcionar una serie de directrices que ayuden a mejorar el uso de los tests por parte de los profesionales, se creó en 1995 la *Comisión de Tests* por el Colegio Oficial de Psicólogos, comisión que funciona de forma coordinada con otras comisiones internacionales creadas con la misma finalidad entre las que merecen destacar la *Task Force* sobre tests de la *Federación Europea de Asociaciones profesionales de Psicólogos (EFPPA)*, o la *International Tests Commission (ITC)*.

8.9. Los tests referidos al criterio frente a los referidos a normas

A finales de los años sesenta y durante la década de los setenta del siglo XX, se había hecho en los EE.UU una fuerte inversión económica en el campo educativo y, lógicamente, se deseaba conocer los resultados de la misma para poder averiguar si se había conseguido que los alumnos alcanzaran unos objetivos mínimos (estándares educativos) y, a la vez, evaluar su nivel de competencia y sus habilidades básicas.

Los tests que se utilizaban habitualmente no se adecuaban bien a los nuevos objetivos planteados, ya que se habían desarrollado para evaluar el grado de aptitud o el nivel de los sujetos en un determinado rasgo, pero interpretando los resultados en relación a los que se habían obtenido en una muestra representativa que había servido para establecer una serie de normas (grupo normativo), sin analizar si los sujetos habían alcanzado o no los objetivos mínimos a nivel educativo: *Tests Referidos a las Normas (TRN)*. Por el contrario, el interés del momento estaba centrado, no en la evaluación del nivel de rasgo o aptitud de los sujetos, sino en la evaluación del rendimiento y en el diagnóstico de las necesidades que pudieran tener los alumnos de cara a una posible intervención. Se trataba de averiguar hasta qué punto los alumnos dominaban el contenido de determinadas materias o eran capaces de resolver correctamente problemas habituales de su vida real, de ahí que se desarrollara otro tipo de tests, *los Tests Referidos al Criterio (TRC)* que iban a tratar de dar solución al problema.

Estos tests proporcionaron las bases para interpretar las puntuaciones de los sujetos en relación a un *dominio* bien definido, más que en relación a un grupo normativo que era como se venía ha-

ciendo, y permitieron a los psicólogos y educadores la evaluación de los estudiantes en función de su conocimiento o no de una materia determinada, en lugar de hacerlo en relación a otros estudiantes (Berk, 1984; Hambleton, 1985).

Hay un acuerdo generalizado en considerar que Glaser (1963) fue el primero que distinguió entre las dos aproximaciones a la medición del rendimiento: la referida a la norma y la referida al criterio y estableció que

...los TRC son aquellos en que los resultados dependen del estatus absoluto de calidad del estudiante, frente a los TRN que dependen del estatus del grupo. (pág. 519).

La definición más aceptada en la actualidad es la de Popham (1978) para quien

...un test referido al criterio se utiliza para evaluar el estatus absoluto del sujeto con respecto a algún dominio de conductas bien definido (pág. 93).

Los trabajos de Glaser suscitaron un enorme interés, y a partir de ese momento aparecieron muchas publicaciones sobre el tema.

Las referencias a trabajos realizados sobre este tipo de tests es continua, sobre todo en algunas revistas como *Journal of Educational Measurement*, *Review of Educational Research*, *Applied Psychological Measurement*, *American Educational Research* y *Journal of Educational Statistics*.

Las diferencias entre los TRN y los TRC no siempre son aparentes a simple vista ya que ambos tipos de tests están formados por ítems de formatos similares, requieren el mismo tipo de operaciones cognitivas de los sujetos, etc. Sin embargo, como recoge Martínez-Arias (1995, pág. 657) hay grandes diferencias tanto en su construcción como en la interpretación de las puntuaciones obtenidas por los sujetos. Estas diferencias hacen referencia a cinco aspectos fundamentalmente:

— *La finalidad de la evaluación:*

Mientras que en los TRN el objetivo es poner de manifiesto las diferencias individuales en la conducta o rasgo que mide el test, en los TRC el objetivo es estimar el rendimiento o conducta del sujeto en los objetivos que mide el test.

— *La construcción del test y la especificación de los contenidos:*

En los TRN a la hora de construir los elementos que formarán el test se suele recurrir a las teorías existentes respecto al rasgo o constructo que se quiere medir, pero no suelen tener una delimitación clara del dominio de contenidos que se van a evaluar. Por el contrario, en los TRC lo primero que hay que hacer es especificar de una manera clara el dominio de contenidos o conductas que se quieren evaluar y el uso que se pretende hacer del test.

— *La forma de seleccionar los ítems:*

En los TRN los ítems deben poner de relieve las diferencias individuales maximizando la varianza del test, y seleccionando ítems de dificultad media y alto poder discriminativo. En los TRC los ítems se seleccionan en función de los objetivos y del uso que se vaya a hacer del test.

— *El significado de las puntuaciones:*

En los TRN la puntuación obtenida por los sujetos se considera un indicador de su puntuación verdadera en un rasgo latente. En los TRC la puntuación es un estimador de la conducta o rendimiento del sujeto en el dominio.

— *La interpretación de las puntuaciones:*

Mientras que en los TRN la puntuación tiene significado únicamente con relación a los resultados del grupo normativo, en los TRC la puntuación tiene significado en términos absolutos.

Estas diferencias implican que a nivel psicométrico se mantenga esta diferenciación entre los dos tipos de tests.

8.10. Los tests adaptativos informatizados (TAI's)

Los avances en el campo de la informática y el hecho de que el ordenador haya bajado su coste de manera ostensible han permitido desarrollar nuevos métodos de enseñanza-aprendizaje y de evaluación del progreso. Asimismo, se han modificado los procedimientos de administración de tests; cada vez son más populares las aplicaciones a través del ordenador. Un ejemplo de este tipo de aplicaciones lo constituyen los *Tests Adaptativos Informatizados (TAI's)*, denominados así porque la selección de los ítems se va haciendo a lo largo del proceso de administración del test, de manera que los ítems se puedan *adaptar* al nivel de cada sujeto y no le resulten ni demasiado fáciles ni demasiado difíciles. Cuando se administran estos tests de forma computerizada, se utiliza el ordenador como medio para hacer la selección de los ítems (Larkin y Weiss, 1974; Olea y col., 1999; Weiss, 1983).

Teniendo en cuenta el interés de la Psicología cognitiva por el estudio de la forma en que las personas procesan y representan mentalmente la información recibida, es fácil darse cuenta que muchos de los procesos medidos tradicionalmente por medio de los tests psicométricos (percepción, memoria, razonamiento, toma de decisión, etc.) entran dentro de su dominio. Este hecho ha sido la causa de que muchos psicómetras se hayan planteado estudiar nuevas formas de construcción de tests de manera que se tengan en cuenta los avances habidos en este terreno.

El lector interesado en conocer las implicaciones que estos desarrollos han tenido para la construcción de tests pueden consultar los trabajos de Embretson (1985, 1993); Messick, (1989); Mislevy (1993);

Pellegrino (1988); Prieto y Delgado (1996); Sternberg (1981, 1988, 1991). Una revisión de los trabajos sobre Inteligencia y Procesamiento de la Información puede encontrarse en Martínez-Arias (1991).

Una exposición más detallada del origen y desarrollo de los tests puede verse en Anastasi (1988), Du Bois (1970), Muñiz, (1998), Nunnally (1970), y Thompson y Sharp (1988) entre otros. Meliá (1990) realiza una aproximación bibliométrica a la Psicometría bastante exhaustiva y una información detallada sobre tests concretos se puede encontrar, entre otros sitios, en las ediciones sucesivas de los *Mental Measurement Yearbooks* editados por Buros que incluyen varios volúmenes con miles de tests incluidos y en las páginas Web de distintas editoriales. Estas publicaciones ponen de manifiesto la gran variedad de tests que se han desarrollado y han convertido a la Psicología en una ciencia positiva.

9. DESARROLLO DE LA TEORÍA DE LOS TESTS

Como consecuencia del auge conseguido por los tests surge la necesidad de desarrollar un marco teórico que sirva de fundamento a las puntuaciones obtenidas por los sujetos cuando se les aplican, posibilite la validación de las interpretaciones e inferencias realizadas a partir de ellas, y permita la estimación de los errores de medida inherentes a todo proceso de medición a través del desarrollo de una serie de modelos.

Partiendo de la base de que las características psicológicas que se intentan medir no son, por lo general, observables de forma directa, y considerando que los tests son los instrumentos de medida que se van a utilizar para llevar a cabo la medición de tales características, cuando se aplica un test a una muestra de sujetos se pueden plantear varias preguntas: ¿en qué medida esas puntuaciones reflejan el nivel de cada sujeto en la característica o rasgo medido?, ¿cómo estar seguros de que la puntuación obtenida es la que le corresponde a cada sujeto realmente? Si en todo proceso de medición, aunque sea en el campo de la física o de la biología, se cometen errores de medida, ¿cuántos más se cometerán cuando se trata de medir variables psicológicas?, ¿qué error está afectando a esas puntuaciones obtenidas al aplicar el test a la muestra de sujetos?

Ante estos problemas es lógico pensar en la necesidad de algún procedimiento que permita analizar, por una parte, la precisión de las medidas obtenidas; es decir, hasta qué punto las puntuaciones obtenidas por los sujetos en un test equivalen a sus puntuaciones verdaderas y cuál es la cuantía del error de medida que las afecta (*fiabilidad de las puntuaciones*) y, por otra parte, la validez de las inferencias o conclusiones que se puedan sacar a partir de las mismas (*validez*).

Para dar respuesta a estos problemas se desarrolló un marco teórico general, la *Teoría de los Tests*, que va a permitir establecer una relación funcional entre las variables observables (a partir de las puntuaciones empíricas obtenidas por los sujetos en los tests o en los ítems que los componen), y las variables inobservables (las puntuaciones verdaderas o el nivel de habilidad de los su-

jetos en el rasgo que se está midiendo) (Gulliksen, 1950; Lewis, 1986 y Lord y Novick, 1968).

Para poder efectuar inferencias a partir de las puntuaciones de los sujetos en los tests, es necesario que la relación entre el nivel de rasgo, constructo o característica, que se quiere medir y las puntuaciones empíricas obtenidas pueda ser establecida a partir de una función matemática o modelo. Cada uno de estos modelos representa un tipo de relación funcional, y mediante una serie de supuestos deberá especificar los factores que influyen en las puntuaciones obtenidas por los sujetos en los tests. En la medida en que los supuestos sean válidos, las inferencias lógicas (matemáticas) que se realicen a partir del modelo describirán de forma correcta las propiedades de las puntuaciones de los tests, en caso contrario estas inferencias serán incorrectas.

Cada modelo podría dar origen a una Teoría de los Tests, pero las que han tenido una mayor incidencia en este campo han sido: la *Teoría Clásica de los Tests* (TCT), y la *Teoría de Respuesta al Ítem* (TRI).

Nota: Dado que en los estudios del Título de Grado de Psicología en la UNED, la Psicometría es una asignatura cuatrimestral, los contenidos de este texto se centrarán en la Teoría Clásica de los Tests. Aquellos que lo deseen podrán completar su formación psicométrica en los estudios de Postgrado.

9.1. La Teoría Clásica de los Tests (TCT)

La TCT se desarrolló, fundamentalmente, a partir de las aportaciones de Galton, Pearson y Spearman, y gira en torno a tres conceptos básicos: las puntuaciones empíricas u observadas (X), las puntuaciones verdaderas (V) y las puntuaciones debidas al error (E). Las primeras corresponden a las puntuaciones obtenidas por los sujetos cuando se les aplica un test, las puntuaciones verdaderas son las que realmente tienen los sujetos en el rasgo o constructo medido y coincidirían con las empíricas en el caso hipotético de que no existieran los errores de medida (E).

Para establecer la relación funcional entre esos tres conceptos la TCT se sirve del modelo más simple, *el modelo lineal*. Este modelo fue desarrollado por Spearman y formulado en una serie de trabajos fechados en 1904, 1907, 1910 y 1913. Para una revisión sistemática de la TCT es necesario acudir a la obra de Guilford (1954) *Psychometric Methods* y sobre todo a la de Gulliksen (1950) *Theory of Mental Tests*.

El modelo lineal de Spearman, cuyos supuestos serán desarrollados en el Tema 4 de este libro, es un modelo aditivo en el que la puntuación observada (variable dependiente) de un sujeto en un test (X) es el resultado de la suma de dos componentes: su puntuación verdadera (variable independiente) (V) y el error (E) que inevitablemente lleva asociado todo proceso de medición. A partir de los supuestos del modelo y de las deducciones que se extraigan a partir de los mismos, se podrá hacer una estimación de esos errores. La expresión formal del modelo es:

$$X = V + E$$

[1.5]

La ejecución de un sujeto al responder a un test en un momento determinado estará afectada por múltiples factores difícilmente controlables, lo que implicará que la puntuación obtenida, su puntuación empírica, no coincida con su puntuación verdadera. Ante la imposibilidad de saber con exactitud cual es esta puntuación verdadera será necesario hacer estimaciones de la misma en base a los supuestos del modelo.

A pesar de que en el modelo lineal de Spearman sólo se encuentra un término de error en el que se incluirían todos los errores aleatorios que están afectando a las puntuaciones empíricas u observadas, los errores pueden provenir de numerosas fuentes: una de las fuentes de error puede ser el mismo sujeto ya que cualquier cosa que le haya ocurrido, su estado emocional, su cansancio y fatiga, etc., puede estar incidiendo en el rendimiento en el test y, por lo tanto, en la puntuación que obtenga; otra fuente de error puede provenir del propio test debido a los ítems que lo forman y al tipo de formato; también las características de los aplicadores del test pueden estar incidiendo en la puntuación de los sujetos en el mismo; otras fuentes pueden ser las condiciones ambientales y las instrucciones que se den, etc.

Aunque resulta imposible separar inequívocamente cuales son los factores que contribuyen a la puntuación verdadera y los que contribuyen al error (Feldt y Brennan, 1989), se han realizado algunos intentos para sistematizar y clasificar el error en función de las posibles fuentes que lo originan (Bock y Wood, 1971; Novick, 1966; Sutcliffe, 1965; Stanley, 1971; Thorndike, 1951, 1989) y se han propuesto algunos modelos que suponen variaciones o extensiones del modelo de Spearman. Las variaciones encontradas entre estos modelos alternativos se deben a distintas matizaciones hechas respecto a los errores. Sin embargo el más ambicioso y global de los intentos realizados para estimar la fiabilidad de un instrumento de medida, analizando de forma sistemática las posibles fuentes de error es el proporcionado por la *Teoría de la Generalizabilidad (TG)* propuesta por Cronbach y sus colaboradores (Glesser, Cronbach y Rajaratnam, 1965; Cronbach, Rajaratnam, Glesser, 1963; Cronbach, Glesser, Nanda y Rajaratnam, 1972) que tiene en cuenta todas las posibles fuentes de error (las debidas a factores individuales, situacionales, características del evaluador y variables instrumentales) e intenta diferenciarlas mediante la aplicación de los procedimientos clásicos de análisis de varianza (AVAR). Las fuentes de variación (factores en términos de ANOVA) se denominan facetas y los niveles de cada factor condiciones. La medida psicológica se convierte en un índice obtenido en una muestra y el problema, como en toda inferencia, será generalizar esa medida. En castellano pueden consultarse Martínez-Arias (1995) o Paz-Caballero (1992).

A pesar del avance que en cierto modo supuso la TG, suele ser considerada más como una extensión de la TCT que como un modelo alternativo. Por otra parte, la complicación de sus diseños y la aparición de nuevos modelos psicométricos englobados bajo la denominación de *Teoría de*

Respuesta al Ítem (TRI) fueron la causa de que su utilización en el marco de la Teoría de los Tests quedara relegada a segundo término.

9.2. Teoría de Respuesta al Ítem (TRI)

Lord (1953) observó que cuando a una muestra de sujetos se les aplicaba un test, o una serie de tests, para evaluar su nivel en un determinado rasgo, la puntuación obtenida dependía del conjunto de ítems o tests utilizados cuando, en realidad, su nivel en el rasgo en el momento de la aplicación no tenía porqué variar. Los sujetos no debían tener puntuaciones altas o bajas en un test en función de que los ítems que lo formaran fueran más fáciles o difíciles. También los estadísticos de los ítems, su índice de dificultad y de discriminación, dependían de la muestra de sujetos utilizada para su cálculo.

Estos dos problemas, junto con el de la indiferenciación del error (sólo había un componente error que englobaba a todos), fueron los que centraron las críticas hechas a la TCT. Los intentos para solucionar el último de ellos ya han sido comentados; para intentar solucionar los dos restantes algunos psicómetras, entre los que se puede citar a Gulliksen (1950) y el mismo Lord (1952, 1953), se interesaron en el desarrollo de teorías y modelos que permitieran describir los niveles de habilidad de los sujetos con independencia de la muestra de ítems o de tareas utilizados para su evaluación, y el cálculo de los estadísticos de los ítems con independencia de la muestra de sujetos utilizada. La solución más adecuada se encontró en el marco de la *Teoría de Respuesta al Ítem (TRI)*, que proporciona una serie de modelos que asumen una relación funcional entre los valores de la variable que miden los ítems (nivel de habilidad de los sujetos en el rasgo medido) y la probabilidad de que los sujetos, en función de su nivel de habilidad, acierten cada ítem. A esta función se la conoce con el nombre de *Curva Característica del Ítem* debido a que, realmente, es la curva que caracteriza a cada uno de ellos. La probabilidad de que un sujeto acierte a cada uno de los ítems no depende ya del propio ítem depende, exclusivamente, del nivel de los sujetos en la variable que mide cada uno de ellos.

En 1952, Lord defendió su tesis doctoral en la que presentó a la TRI como un modelo o teoría con entidad propia, de ahí que sea considerado el padre y fundador de la TRI. Como resultado de su tesis se publicó en *Psychometric Monographs* nº 7, una monografía bajo el título *A Theory of Test Scores*, a este trabajo siguieron otros que marcaron el comienzo de una nueva manera de trabajar en el campo de la Psicometría (Birnbaum, 1957, 1958a, 1958b; Lord y Novick, 1968; Rasch, 1960).

El desarrollo de estos modelos supuso un gran avance en la Teoría de los Tests; sin embargo, la dificultad de utilizarlos en la práctica sin la ayuda de los ordenadores fue la causa de que su gran desarrollo no llegara hasta finales del siglo XX, cuando ya el uso de los ordenadores personales fue habitual y asequible para una gran mayoría y, además, se desarrollaron los programas de software necesarios para su utilización.

A pesar del gran desarrollo de la TRI hacia 1980, la TCT sigue en auge ya que hay problemas que se pueden solucionar más eficaz y rápidamente dentro de este marco.

Nota: El lector interesado en la Teoría de Respuesta al Ítem puede consultar, en castellano, los siguientes textos introductorios: Martínez-Árias, M.R. (1995) *Psicometría: Teoría de los tests psicológicos y educativos*. Madrid: Síntesis (capítulos 10 y 11); Muñiz, J. (1997). *Introducción a la Teoría de Respuesta a los Ítems*. Madrid: Pirámide; Santisteban, C. (1990/1995). *Psicometría. Teoría y práctica en la construcción de tests*. Madrid: Norma (capítulos 10 y 14).

10. EJERCICIOS DE AUTOEVALUACIÓN

A continuación se proponen una serie de afirmaciones que pueden ser correctas o incorrectas; el lector, después de leerlas detenidamente, deberá responder en un sentido o en otro razonando su respuesta.

1. La Teoría de la Medición es uno de los campos de la Psicometría.
2. Las dos vías a través de las cuales se desarrolló la Psicometría fueron los estudios de Psico-física y las investigaciones acerca de las diferencias individuales.
3. Los métodos psicofísicos se desarrollaron para su utilización en el campo de la percepción.
4. A través de los métodos de escalamiento psicológico se pueden medir variables que no tengan ninguna dimensión física subyacente.
5. Los métodos psicofísicos están vinculados a la Psicología correlacional.
6. Los tests de Galton medían funciones mentales.
7. Los atributos psicológicos son variables directamente observables.
8. Los métodos de escalamiento psicofísico pretenden ordenar a los sujetos a lo largo de un continuo psicológico.
9. Cuando hay varias dimensiones subyacentes a los estímulos, éstos reciben un valor escalar en cada una de ellas.
10. La Psicometría utiliza como método de investigación el método científico.
11. Podemos considerar a los tests como instrumentos de medición.
12. Spearman desarrolló un modelo para las puntuaciones de los tests.
13. Un problema fundamental en la medición psicológica es el del control del error de medida.
14. El cociente intelectual es una norma cronológica que permite la ordenación de los sujetos.
15. Binet fue el primero en considerar la importancia de los procesos mentales superiores en el estudio de las diferencias individuales.
16. Las puntuaciones obtenidas por los sujetos en un test referido al criterio se comparan con las obtenidas por un grupo normativo.
17. El objetivo de los tests referidos a las normas es poner de manifiesto las diferencias individuales en el rasgo que miden.
18. Los Tests Adaptativos Informatizados (TAI s) son los mismos tests de papel y lápiz pero aplicados por ordenador.

-
19. Una de las críticas a la Teoría Clásica de los Tests es que los parámetros de los ítems dependen de la muestra de sujetos a los que se les aplican y las puntuaciones de los sujetos en el rasgo medido dependen de los ítems a los que responden.
 20. Los *métodos directos* de elaboración de escalas psicofísicas utilizan una escala de respuesta.
 21. Los *métodos directos* se utilizan para la obtención de umbrales.
 22. El umbral absoluto marca el origen de la escala de sensación.
 23. El paso de la sensación a la no sensación viene determinado por el umbral diferencial.
 24. En el modelo escalar de Thurstone los estímulos se ordenan a lo largo de un continuo físico.

11. SOLUCIONES A LOS EJERCICIOS DE AUTOEVALUACIÓN

1. La afirmación es correcta.

Hemos considerado a la Psicometría como una disciplina, dentro del área de la Psicología, que tiene unas funciones concretas, tanto en lo que se refiere a sus implicaciones teóricas como prácticas. En este sentido, la Teoría de la Medición tiene como objetivo legitimar y justificar la medición de variables psicológicas dándole una fundamentación teórica.

2. La afirmación es correcta.

Como hemos comentado anteriormente, las dos vías de desarrollo de la Psicometría fueron los estudios de Psicofísica y las investigaciones acerca de las diferencias individuales. Aunque en un principio pudiera parecer que ambas vías de desarrollo siguieron caminos distintos, podemos comprobar la incidencia que tienen los avances logrados en cada una de ellas sobre la otra.

3. La afirmación es correcta.

Los estudios de Psicofísica tratan de relacionar la magnitud de los estímulos con la percepción que de ellos tienen los sujetos; es decir, con la sensación que les producen.

4. La afirmación es correcta

Así como los métodos psicofísicos se utilizan para estudiar las relaciones entre un conjunto de estímulos que varían a lo largo de un continuo físico y las sensaciones que producen en los sujetos que varían a lo largo de un continuo psicológico, como ocurre, por ejemplo, si queremos establecer una correspondencia entre la intensidad de luz de una serie de estímulos y la sensación de luminosidad que producen; si tratamos de medir las sensaciones que producen en una muestra de sujetos una serie de obras de arte, la carga de violencia de determinadas escenas de películas, etc., nos encontramos con que los estímulos no varían a lo largo de ningún continuo físico sino psicológico y las sensaciones que producen en los sujetos también varían a lo largo de un continuo psicológico. Por eso, para la medición de este tipo de variables psicológicas se utilizan otros métodos de escalamiento, los métodos psicológicos, cuyo principal impulsor fue Thurstone.

5. La afirmación es incorrecta.

Los métodos psicofísicos están vinculados a la Psicología experimental; son los estudios de las diferencias individuales los que están vinculados a la Psicología correlacional.

6. La afirmación es incorrecta.

Los tests de Galton, que pueden ser considerados los primeros tests psicométricos, tenían un marcado carácter sensorial y motor. Aunque Galton pensaba que los datos antropométricos que estaba reuniendo para sus investigaciones le podían valer para estudiar las dimen-

siones de la mente del hombre, al final de sus días debió admitir que esas medidas no tenían valor como medidas de la inteligencia.

7. La afirmación es incorrecta.

Las variables psicológicas o atributos psicológicos, son variables hipotéticas denominadas constructos que no se pueden medir de forma directa porque no son variables directamente observables. Es preciso determinar una muestra de conductas representativas del atributo que queremos estudiar y, dado que estas conductas son variables observables susceptibles de medición, a partir de ellas haremos inferencias acerca del atributo.

8. La afirmación es incorrecta.

Por una parte los métodos psicofísicos no ordenan a los sujetos puesto que son métodos para escalar (ordenar) estímulos y, por otra, los métodos de escalamiento psicofísico permiten relacionar dos continuos, un continuo físico a lo largo del cual varían los estímulos y un continuo psicológico a lo largo del que se sitúan las sensaciones que producen los estímulos.

9. La afirmación es correcta.

A diferencia de los métodos de escalamiento unidimensional en los que los estímulos reciben un valor escalar en la única dimensión que se mide, en los métodos de escalamiento multidimensional, al estar considerándose varias dimensiones a la vez, los estímulos recibirán un valor escalar en cada una de ellas.

10. La afirmación es correcta.

La Psicología, como disciplina científica que es, utiliza el método científico para la adquisición del conocimiento, adaptándole a sus necesidades particulares.

11. La afirmación es correcta.

El método de los tests se desarrolla para el estudio de las diferencias individuales y, para ello, desarrolla los tests como instrumentos de medición.

12. La afirmación es correcta.

En efecto, Spearman desarrolló un modelo lineal que supuso un gran desarrollo para el método de los tests (Teoría Clásica de los Tests). Este modelo partía del supuesto de que la puntuación observada a través de los tests (X) está formada por dos componentes, una componente debida a la verdadera medida del rasgo (puntuación verdadera del sujeto en el rasgo) y otra una componente errónea (el error asociado a todo proceso de medición), y que la relación entre estas dos componentes es aditiva.

13. La afirmación es correcta.

En todo proceso de medición, aún en los llevados a cabo en las ciencias exactas, se cometen errores de medición. Este problema se agrava cuando las características que se quieren medir

no pueden ser observadas directamente y han de ser medidas mediante indicadores. En este caso el control del error cometido es un problema fundamental.

14. La afirmación es correcta.

Se trata de un término acuñado por Stern (1912) y que equivale al cociente entre la edad mental y la edad cronológica, multiplicado por cien para evitar los decimales.

15. La afirmación es correcta.

Binet comprendió que las sensaciones no jugaban un papel demasiado importante en la Psicología diferencial y que había que recurrir al estudio de los procesos mentales superiores.

16. La afirmación es incorrecta.

Una de las diferencias entre el enfoque normativo y el de los tests referidos al criterio es precisamente que, en estos últimos, no se necesita un grupo normativo para la interpretación de las puntuaciones obtenidas por los sujetos, sino que éstas se interpretan en relación a un dominio de contenidos o conductas.

17. La afirmación es correcta.

Los resultados se interpretan en relación a los obtenidos por el grupo normativo.

18. La afirmación es incorrecta.

Una cosa son los tests aplicados por ordenador y otra los tests adaptativos informatizados. En éstos, los sujetos no tienen que contestar ni a los mismos ítems ni a todos los ítems de un tests. La selección de cada ítem se va haciendo de manera que se vayan *adaptando* al nivel de cada sujeto.

19. La afirmación es correcta.

En el marco de la TCT, los valores de los parámetros de los ítems dependen de la muestra de sujetos a los que se les han aplicado. Si el índice de dificultad de un ítem se obtiene calculando la proporción de sujetos que han acertado ese ítem, es fácil darse cuenta de que esa proporción variará en función del nivel de los sujetos. Por otra parte, el nivel de aptitud o habilidad de los sujetos depende de que los ítems a los que respondan sean más fáciles o difíciles. Este problema no tenía una solución real dentro del marco de la TCT y hubo que esperar al desarrollo de la TRI para que se pudiera solucionar.

20. La afirmación es correcta.

En eso se diferencian de los métodos indirectos asociados a la psicofísica de Fechner, ya que éstos utilizan una escala de sensación elaborada a base de ir sumando las diferencias apenas perceptibles (dap).

21. La afirmación es incorrecta.

Los métodos directos no implican el cálculo de umbrales, en estos métodos el sujeto emite de forma directa su respuesta.

22. La afirmación es correcta.

El umbral absoluto es el valor mínimo que tiene que tener un estímulo para poder ser percibido por el sujeto. Este valor mínimo en la escala física se empareja con el valor cero de la escala de sensación y, por lo tanto, marca su origen.

23. La afirmación es incorrecta.

El paso de la sensación a la no sensación (o viceversa) equivale, en la escala física, al umbral absoluto. El umbral diferencial es el incremento mínimo que tiene que experimentar la magnitud de un estímulo para que el sujeto perciba que ha habido un cambio.

24. La afirmación es incorrecta.

Precisamente la gran aportación de Thurstone fue elaborar un modelo de escalamiento en el que no fuera necesario recurrir a ningún continuo físico.

12. BIBLIOGRAFÍA COMPLEMENTARIA

Barbero, M.I. (2007). *Métodos de elaboración de escalas*. Madrid: UNED.

A lo largo del libro los alumnos podrán encontrar información sobre algunos de los principales métodos de escalamiento

Muñiz, J. (1998, 2008). *Teoría Clásica de los tests*. Madrid: Pirámide

En el primer capítulo se hace una buena introducción sobre el origen y desarrollo de los tests y de la teoría de los tests.