

1.1. Aproximación al concepto de Psicometría

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)

Después de esta definición, se puede decir que la Psicometría deberá ocuparse de:

- La **justificación y legitimación de la medición psicológica**
- Las **implicaciones prácticas y aplicadas que dicha medición conlleva**

· 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
- c) Establecer las condiciones que permitan llevar a cabo el proceso de medición.

· Las **implicaciones prácticas y aplicadas que dicha medición conlleva:**

- a) Proporcionando los métodos necesarios que indiquen, en cada caso concreto, cómo se debe llevar a cabo la cuantificación
- b) Construyendo los instrumentos necesarios y adecuados para poder efectuarla.

1.2. La Psicometría en el marco de la Metodología de la Psicología como ciencia del Comportamiento

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.

La Psicometría, tiene como objeto de estudio la conducta humana y como el método, el método científico. 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.

1.3. La medición psicológica

La mayoría de los atributos psicológicos como por ejemplo la inteligencia, el autoritarismo y la introversión, son conceptos abstractos, denominados *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. Se consideró que se trataba de un *proceso mediante el cual se enlazan conceptos abstractos (los constructos inobservables directamente)*, con indicadores empíricos observables directamente (las conductas).

Así, 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 y, por lo tanto, pueden ser medidas mediante el instrumento de medida 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*, 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.

A parte de todo esto, nos encontramos con otro problema, las unidades de medida. En Psicología podemos abordar el problema de dos formas:

- **Referida a normas:** Consiste en comparar los resultados obtenidos por el sujeto con los obtenidos por un grupo de sujetos que forman el llamado *grupo normativo* y que pertenecen a su misma población.
- **Referida al criterio:** Los resultados obtenidos se comparan con un criterio (un punto crítico) y la superación o no del mismo es lo que va a dar significado a las puntuaciones obtenidas.

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:

1. Estimar los errores aleatorios que conlleva toda medición (fiabilidad de las medidas)
2. Garantizar que la misma no es algo inútil, sino que sirve para explicar y predecir los fenómenos de interés (validez de las medidas)

1.4. Orígenes y vías de desarrollo de la Psicometría

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 Test (modelos), que posibilitan la asignación de valores numéricos a los sujetos y, por lo tanto, al *escalamiento de los sujetos*.

1.5. 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 con unas normas o reglas.

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 pueden situarse de forma ordenada a lo largo de ese continuo.



1.6. Desarrollo de los métodos de escalamiento de estímulos

Podemos clasificar los métodos de escalamiento en:

- Método de escalamiento psicofísico (no entra en el temario)
- Método de escalamiento Psicológico (se desarrolla a continuación)

1.6.1. Método 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.

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 sujetos 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 diferencias 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. Thrustone desarrolló un modelo matemático cuyas ecuaciones se conocen con el nombre de *Ley del Juicio Comparativo* que utiliza el *Método de las Comparaciones Binarias*. Y posteriormente desarrollaría otro modelo cuyas ecuaciones se conocen como la *Ley del Juicio Categórico* que utiliza el *Método de los Intervalos Sucesivos*, el *Método de los Intervalos Aparentemente iguales* y el *Método de Ordenación de Rangos*.



Guttman desarrolló un nuevo modelo para el escalamiento conjunto de sujetos y estímulos. 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 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.

1.7. Orígenes y desarrollo de los tests

Un tests 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 existente 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
- La interpretación que Spearman hace de ella, considerando que la correlación entre dos variables indica que ambas tienen un factor común.

1.7.1. La importancia de los trabajos de Galton

Galton 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.

Fue el primero en aplicar el concepto estadístico de distribución normal, de media, de mediana, varianza y correlación a datos psicológicos.

Para poder llevar a cabo sus mediciones, Galton construyó y utilizó una serie de instrumentos que pueden ser considerados los primeros *Test 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.



1.7.2. Los primeros test mentales

James McKeen Cattell en contacto con Galton utilizó por primera vez el término *Test Mental*. Tenía un carácter sensorial y motor fundamentalmente, ya que incluían medidas de energía muscular, velocidad de movimiento, 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.

1.7.3. Desarrollo de los primeros tests de inteligencia

Alfred Binet 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.

Binet trabajó con Simon y realizaron una serie de investigaciones cuyo resultado fue la publicación de la primera escala de inteligencia. A diferencia de los tests de Galton, este test ya no se ocupaba de las funciones motoras y sensoriales sino de la capacidad de comprensión y razonamiento de los sujetos. Esta escala es conocida como la *Escala de Binet-Simon*.

Terman 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. Con esto se establece la primera escala de puntuaciones referida a las aptitudes mentales.

$$CI = \frac{Edad\ mental}{Edad\ cronológica} \cdot 100$$



1.7.4. Los tests colectivos

Todos los tests desarrollados hasta el momento eran de aplicación individual. En la Primera Guerra Mundial, se desarrollaron los primeros tests colectivos de inteligencia. Los *Tests Alpha* y *Beta*. El *Tests Alpha* para la población general y el *Tests Beta* para los reclutas analfabetos o aquellos que no dominaban el inglés.

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*.

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.

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.

1.7.5. Las baterías de aptitud múltiple

Cuando la técnica del análisis factorial empezaba a dar sus frutos, comenzaron a desarrollarse las *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.

Entre las primeras baterías merece destacar la *Batería de Aptitudes Mentales Primarias* de Thurstone 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.



1.7.6. Los tests de personalidad

Se pueden considerar a Kraepelin como un precursor de los *Tests de personalidad*, en 1894 utilizó el *Test de Asociación Libre* con pacientes normales par 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.

El prototipo de cuestionario de personalidad con carácter psicométrico es la *Hoja de Datos Personales* de Woodworth 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.

De los primeros tests objetivos de personalidad hay tres que han sobrevivido en la actualidad:

- Una serie desarrollada por Guilford y sus colaboradores
- Una serie desarrollada por Cattell y sus colaboradores
- El *Inventario Multifásico de Personalidad de Minnesota* de Hathaway y Mckinley

1.7.7. 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) y la *Escala de Preferencia de Kuder*.

En cuando a la medida de las actitudes, creencias y opiniones, su desarrollo 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*.

Algunos años más tarde Likert 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 que siguen la *Técnica de Likert* están basadas en los mismos principios que la mayoría de los test de aptitudes.

Con el fin de medir el significado connotativo, también llamado significado afectivo o subjetivo, que determinados estímulos tienen para los sujetos, Osgood 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*.

1.7.8. 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 test, pero también un cierto abuso, lo que condujo a numerosas críticas sociales e incluso a la prohibición de su uso en determinados estados.

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

Glaser 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.

La definición más aceptada en la actualidad es la de Popham:

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

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 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:

1. La finalidad de la evaluación

2. La construcción del test y la especificación de los contenidos

3. La forma de seleccionar los ítems

4. El significado de las puntuaciones

5. La interpretación de las puntuaciones

1. La finalidad de la evaluación:

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

2. La construcción del test y la especificación de los contenidos:

- 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.

- 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

3. La forma de seleccionar los ítems

- 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.
- TRC: Los ítems se seleccionan en función de los objetivos y del uso que se vaya a hacer del test.

4. El significado de las puntuaciones

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

5. La interpretación de las puntuaciones

- TRN: La puntuación tiene un significado únicamente con relación a los resultados del grupo normativo
- TRC: La puntuación tiene significado en términos absolutos.

1.7.10. Los test adaptativos informatizados (TAI's)

Los avances en el campo de la informática 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.

1.8. Desarrollo de la Teoría de los Tests

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 afectado a estas puntuaciones obtenidas al aplicar el test a la muestra de sujetos?

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 y las inobservables.



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.

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).

1.8.1. La Teoría Clásica de los Test (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) que son las obtenidas por los sujetos cuando se les aplica un test
- Las puntuaciones verdaderas (V) son las que realmente tienen los sujetos en el rasgo o constructo medido que coincidirán en el hipotético caso de que no existieran los errores de medida.
- Las puntuaciones debidas a 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*.

$$X = V + E$$



1.8.2. La Teoría de Respuesta al ítem (TRI)

Lord 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 por qué variar. Los sujetos no debían tener puntuaciones altas o bajas en un test en función de que los ítems que los 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 fueron los que centraron las críticas hechas a la TCT. La solución más adecuada para los dos primeros 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*. 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.