

Principios de psicometría aplicados a la evaluación del dolor

C. LÁZARO¹, J.E. BAÑOS² Y R. TORRUBIA³

«Es perfectamente posible que muchas de nuestras teorías sean de hecho certeras pero, incluso si son verdaderas, nunca podemos saberlo con certeza»

K. Popper

RESUMEN

Dada la subjetividad implícita al dolor, es difícil obtener métodos que obvien la opinión del paciente. Por ello, es indispensable desarrollar métodos que tengan en cuenta su opinión, pero que a la vez sean válidos y fiables. La aplicación de los principios psicométricos al desarrollo de instrumentos para evaluar el dolor constituye una inevitable obligación para que éstos puedan ser aplicados adecuadamente en la práctica clínica. En el presente artículo se realiza una revisión de los conceptos más relevantes, con especial hincapié en los paradigmas para la valoración de los métodos psicométricos, como fiabilidad, validez y análisis factorial.

Palabras clave: Métodos psicométricos: fiabilidad, validez, análisis factorial.

ABSTRACT

Given the implicit subjectivity of pain, the building of evaluation tools that avoid the opinion of patient is difficult. Therefore, it seems indispensable to develop methods that consider their opinion and still were valid and reliable. The application of psychometric principles to development of instruments devoted to evaluate pain is compulsory to permit its use in the clinical setting. The present article reviews the most relevant concepts related with psychometric development, and outlines the main paradigms used to establish the usefulness of these instruments, as reliability, validity and factor analysis. (DOLOR 2007;22:12-25)

Corresponding author: Rafael Torrubia, rafael.torrubia@uab.es

Key words: Psychometric methods: reliability, validity, factor analysis.

¹Departamento de Farmacología, Terapéutica y Toxicología
Universitat Autònoma de Barcelona

Bellaterra, Barcelona
²Departamento de Ciencias Experimentales y de la Salud
Universitat Pompeu Fabra
Barcelona

³Departamento de Psiquiatría y Medicina Legal
Universitat Autònoma de Barcelona
Bellaterra, Barcelona

Dirección para correspondencia:

Rafael Torrubia
Departamento de Psiquiatría y Medicina Legal
Facultad de Medicina
Universitat Autònoma de Barcelona
08193 Bellaterra
E-mail: rafael.torrubia@uab.es

ATRIBUTOS, TEORÍAS Y MEDICIONES

El método científico moderno desarrollado por C. Bernard rompió con el enfoque tradicional de la ciencia basado en aproximaciones empíricas, abriendo paso al conocimiento basado en evidencias contrastadas y repetibles que hoy conocemos. Recordemos el aserto del filósofo positivista K. Popper: «Una teoría científica es un modelo matemático que describe y codifica las observaciones: una teoría es adecuada cuando describe una gama de fenómenos a partir de algunos postulados simples y permite efectuar predicciones susceptibles de ser analizadas». Las palabras «modelo matemático» y «análisis» dan el sentido que permite diferenciar la ciencia de las formulaciones empíricas. No es pretencioso decir que la medición de fenómenos es la base sobre la que se sustenta el desarrollo de la teoría científica.

En el estudio de un fenómeno, el observador es un validador externo y su función es proveerse de procedimientos que representen el sistema que pretende medir, de forma que pueda registrar de forma sistemática los fenómenos observables. Éstos son un conjunto de atributos, definidos como una característica particular del objeto. Conviene, además, que esta representación permita conocer qué cantidad de atributo está presente en dicho objeto.

La medición es el proceso mediante el cual se asignan símbolos a objetos teniendo como base unas reglas establecidas de antemano, con el fin de obtener una representación de alguno de sus atributos (escalas), a la vez que pretende definir si los objetos caen dentro de la misma categoría o diferente respecto de un atributo (clasificación). Existe, por lo tanto, una conversión de los hechos observables hacia el mundo abstracto de símbolos que los representan: medición implica abstracción.

La relación entre la medida y el objeto puede estar basada en la aproximación o bien en la convención, no teniendo por qué ser mejor un sistema que otro. Recordemos que el término «mejor» ya implica una valoración por parte de algún observador y una comparación entre sistemas teniendo como base, otra vez, unas reglas. La labor de creación de escalas consiste en relacionar constructos que no pueden ser directamente observados, pero de los que se conoce su existencia, con indicadores de las manifestaciones del fenómeno.

Existen muchas formas de generar instrumentos de medición, pero, para proveer de rigor científico este

proceso, se requiere la labor de estandarización. Este proceso asegura que el instrumento: 1) contenga reglas claras; 2) su aplicación sea práctica; 3) no requiera habilidades especiales por parte de los entrevistadores, y 4) sus resultados no dependan del administrador¹. Estos conceptos están siendo objeto de continuo debate por su contenido abstracto (aunque todavía podemos definir de forma razonable lo que son «reglas claras» o lo que son «habilidades especiales»).

LAS DISTINTAS ESCUELAS DE MEDICIÓN

Los sistemas de medición y sus unidades acompañan el devenir de las corrientes del pensamiento. Es interesante abordar los distintos enfoques teóricos que se han aplicado al proceso de la medición.

Antropocentrismo

En su origen las cosas están hechas a conveniencia del hombre y nadie se plantea que puedan ser de otra forma; las necesidades de medición pasan por tener una utilidad estrecha en relación con la función social o la técnica laboral de su época. Las unidades se ajustan a las necesidades concretas y son a menudo antropomórficas (el palmo, el paso, la braza), preceden a los sistemas métrico (metro, gramo, litro) e imperial británico (pulgadas, libras, pintas) establecidos por convenio, pero, dado que el pueblo llano no se daba por aludido, coexisten. La cuestión se complica con la revolución industrial y una nueva explosión de unidades de medición, basadas ahora además en la profesión. El peso es medido de forma distinta por ingenieros o por joyeros, dejando aparte las utilizadas por marinos². Muchas de estas unidades están vigentes y hay que reconocer que son prácticas y suelen facilitar las cosas a nivel local.

Operacionalismo

La Revolución francesa trae nuevas ideas igualitarias, entre ellas una ley promulgada por la Asamblea Nacional francesa para crear un estándar de pesos y medidas que evite ventajas o disputas entre naciones. Por desgracia, la *Royal Society of London* no respondió a la invitación de la *Academie Française des Sciences* prevista para armonizar los sistemas de medición³.

A pesar de la falta de convención, se pueden aplicar en este terreno tablas de conversión (v. g. kilómetros a millas). La utilización de estas unidades no trae asociada una nueva filosofía del proceso de medición. Cualquier concepto teórico debe ser equivalente, o reducible, a una operación que lo defina. Para P. Duhem o P. Bridgman⁴, «longitud» significa que el objeto de medición se contrasta con un instrumento convenido, como en el caso de la longitud utilizando el metro y nada más: «la medición, en su sentido más general, puede ser tomada como la descripción mediante el uso de números». Una medición singular, obtenida mediante una operación específica, es una válida y suficiente caracterización del objeto, evento o proceso que se mide.

En el contexto de la medición de hallazgos biomédicos, cabe admitir que la cuestión se torna más compleja. La filosofía multidisciplinaria y la óptica multidimensional actual no pueden reducirse a la utilización de un sistema único y establecido de forma arbitraria que limita el espectro de atributos sometidos a estudio y resulta inadecuada para determinar la magnitud del objeto que se estudia.

Representacionalismo

En la reunión anual de la *British Association for the Advancement of Sciences* de 1874, el físico irlandés G.J. Stoney, en contraposición a la arbitrariedad de las unidades de medición de los operacionalistas, aboga por unidades fundamentales de la física. Esta sugerencia tiene como antecedente la idea del científico francés J. Babinet de utilizar la longitud de onda de la luz para definir la longitud en 1827, si bien no existían los medios para realizar estos cálculos. Estas unidades son admitidas rápidamente por algunos científicos y, por ejemplo, M. Planck las describe como unidades naturales³. Estas unidades suelen ser demasiado grandes o demasiado pequeñas para nuestro normal desenvolvimiento.

Para los representacionalistas, el trabajo del científico es descubrir fenómenos y la relación entre los fenómenos que existen en la naturaleza, caracterizándolos mediante expresiones matemáticas formales. En este caso, el científico debe asignar los números de forma muy cuidadosa para que sirvan como fieles representaciones de las verdaderas relaciones subyacentes entre fenómenos que ocurren en la naturaleza.

El representacionalismo no se preocupa de las construcciones hipotéticas: el fenómeno, observable o

no, no nace de la imaginación del teórico; el científico debe descubrir, no construir modelos o esquemas teóricos que se aproximen a la realidad.

Algunos argumentos en contra de esta corriente se basan en el hecho de que, incluso siendo cuidadosos a la hora de representar la realidad, nuestra visión siempre estará inferida por el mismo proceso de medición o, en el caso de estudiar una experiencia como el dolor, por la configuración del estado del sistema nervioso determinado por su estructura evolutiva; en otras palabras, utilizamos el instrumento de análisis para estudiar el instrumento de análisis⁵. Por otra parte, son numerosas las teorías científicas que han podido predecir la existencia y los atributos de constructos previamente a su observación, basándose en circunstancias indirectas (incluso la existencia del electrón por el mismo G.J. Stoney).

Clasicismo

Para los clasicistas la ciencia implica una estructura definible a la que podemos llamar teoría, o bien una serie de conceptos y asunciones mantenidas en común por una comunidad de investigadores que determina el significado de una medida. Estos conceptos y asunciones están sujetos a modificaciones de la teoría.

En la estimación están involucrados tres procesos: a) el procedimiento experimental que produce datos; b) un modelo que relaciona los datos a los parámetros que están siendo estimados, y c) operaciones matemáticas que determinan los valores de los parámetros que están siendo estimados por los datos⁶.

Así, la ciencia debiera caracterizarse por pocos parámetros y muchos modelos para posibilitar mediciones independientes por diferentes métodos que sirvan para confirmar la validez de cada uno. Los principios del clasicismo serían los siguientes:

- La medición depende del contexto y es inseparable de la teoría científica.
- Los atributos mesurables pueden existir en la naturaleza (variables manifiestas) o en la teoría (variables latentes).
- Ninguna medida aislada será un perfecto indicador de una variable latente.

Los clasicistas crean modelos de la realidad más que trabajar con ella directamente, lo que implica una sucesión de aproximaciones hasta adecuar la teoría a la realidad.

IMPORTANCIA DE LA MEDICIÓN

Los motivos por los que algunos investigadores se toman la molestia de crear instrumentos de medición estandarizados se pueden enumerar bajo los paradigmas que siguen:

Cuantificación de atributos

Una forma básica de medir un atributo basada en juicios personales es superada por un método estandarizado al ser éste capaz de proveer de indicativos más sutiles. Esta diferencia lleva a la posibilidad de utilizar métodos más potentes, como el análisis matemático y bioestadístico, que permitirán un conocimiento más profundo de los resultados experimentales y la formulación de hipótesis.

Clasificación

Una medición puede determinar si los objetos caen en determinadas categorías (previamente definidas) respecto a un atributo. Por ejemplo, las cifras de glucemia de una curva obtenidas de un individuo y presencia en él de diabetes *mellitus*, o bien rasgos psicológicos presentes en un individuo y neuroticismo.

Sistematización e intercambio de conocimientos

Una de las bases del método científico es que los resultados obtenidos por un investigador puedan ser verificables de forma independiente por otros. Para que esto ocurra, debe haber un método de cuantificación admitido por todos ellos, incluso contemplando la posibilidad de que no sea un sistema perfecto pero del que todos conocen sus limitaciones. Una vez obtenido un instrumento estandarizado, diversos investigadores pueden trabajar en lugares distantes e intercambiar información sobre los resultados obtenidos con la seguridad de que las puntuaciones en todos los lugares representan un mismo constructo y contienen parámetros similares del proceso que se intenta medir. Existe la salvedad de las diferencias entre grupos culturales, étnicos y sociales que, a menudo, requiere realizar validaciones de estos instrumentos, como se analizará más adelante.

Generalización y formulación de teorías

La ciencia intenta dar una explicación a una serie de fenómenos observables mediante unos principios ra-

zonados (leyes). Las teorías son las construcciones hipotéticas en las que, asumiendo una forma de funcionamiento de los hechos observables, se intenta dar explicación plausible a su forma de operar. Estas leyes intentan ser generalizables y la cuantificación ayuda a realizar estimaciones de tipo estadístico en las que se refleje la posibilidad de que ocurra un evento a partir de unos determinados supuestos. Por otra parte, es una aspiración común a cualquier disciplina el verificar sus resultados con los obtenidos por otra en el estudio de un mismo fenómeno. Las técnicas psicométricas permiten estimar las relaciones entre distintos hallazgos, y este hecho ayuda a romper las barreras existentes entre los distintos enfoques, creando una sensación de unificación de ideas sobre el fenómeno que se estudia y permitiendo formular o refutar axiomas, paso previo a la formulación de hipótesis. En algunas ocasiones, los instrumentos de medición pueden tener la aspiración de reflejar un modelo científico determinado. Aquí, la medición relaciona la hipótesis con variables latentes (un fenómeno que no puede ser directamente observado). Los cálculos que los instrumentos psicométricos permiten realizar pueden dar luz sobre si la teoría da una explicación a los hallazgos realizados, pero conviene tener cuidado con esta situación ya que a menudo la teoría nos puede estar marcando lo que vamos a observar, no la realidad del universo observable.

Ahorro de tiempo, esfuerzo y recursos

El desarrollo de sistemas de cuantificación puede devenir un arduo trabajo, si bien, una vez obtenidos, el trabajo de unos pocos permite al cuerpo de profesionales e investigadores dedicar todo el tiempo que tendrían que estar realizando evaluaciones subjetivas a otras cuestiones más creativas y, de este modo, favorece el progreso de la ciencia.

CONDENSACIÓN DE LA TEORÍA PSICOMÉTRICA

Los métodos utilizados en los primeros pasos de la psicometría han encontrado apoyo en la ciencia psicofísica. Esta rama de la psicología estudia la relación entre las propiedades físicas de los estímulos y las experiencias sensoriales en humanos y de comportamiento en otras especies. El ejemplo más cercano de valoración psicofísica es el que realizan los audiólogos durante un examen audiométrico al pre-

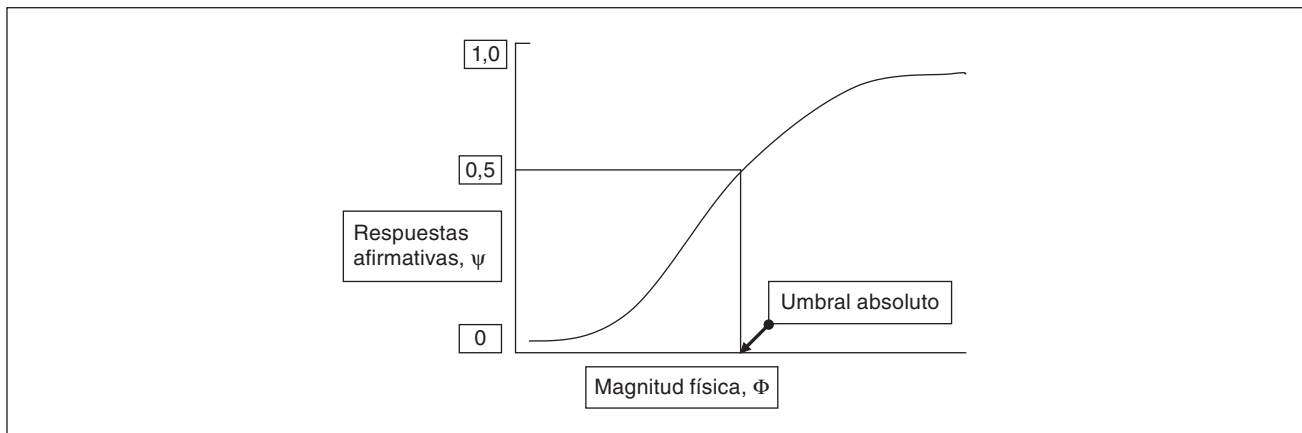


Figura 1. Modelo de Fechner.

sentar al individuo, en una cámara anecoica, un estímulo auditivo y determinar el momento en que percibe sonido utilizando distintas frecuencias; la aplicación de este proceso a aspectos psicológicos de un individuo será la psicometría.

Las bases de esta disciplina las encontramos en los estudios de G.T. Fechner, que intenta sistematizar las relaciones entre la variación en los estímulos físicos (representadas por Φ) y las experiencias subjetivas (representadas por Ψ). En el caso de la medición de características psicológicas (inteligencia, psicopatología) no hablaremos de psicofísica, ya que el atributo a medir no es una dimensión física. No obstante, la psicofísica se ha provisto de metodología utilizada posteriormente en psicometría, y por ello se mencionan aquí, aunque sea someramente, algunos de sus fundamentos.

G.T. Fechner utilizó el método de adaptación para determinar el valor conocido como umbral absoluto (patrón a partir del cual empieza a percibirse un estímulo), y el umbral diferencial (patrón a partir del cual se reconoce variación en el estímulo), que representa la unidad basándose en la diferencia apenas perceptible (DAP) entre estímulos.

Si representamos gráficamente en dos ejes las magnitudes físicas en abscisas (Φ) y la proporción de respuestas afirmativas en ordenadas (Ψ), habitualmente obtendremos una curva de respuestas de forma en S u ojiva; la curva obtenida para un sujeto particular se situará a la izquierda si tiene una capacidad mayor para responder al estímulo, y viceversa. De acuerdo con la ley de Fechner, asumiendo intervalos separados por una DAP, se definen unidades iguales en una escala de sensación que se pueden expresar matemáticamente mediante la conocida ecuación de Fechner:

$$\Psi = b \log (\Phi) + a$$

donde b y a son constantes de escala (habitualmente el valor a se elige de forma que haga que Ψ sea 0 cuando Φ está en el umbral). Esta ecuación permite obtener gráficas como la que se representa en la figura 1. La función logarítmica proporciona la ventaja de que razones físicas iguales producen diferencias sensoriales iguales.

El modelo de Fechner asume que el estímulo alcanza un cerebro quiescente en situación de espera pasiva de un estímulo que llega prácticamente sin modulación a lo largo de las vías que conducen la sensación; sin embargo, es conocido que el sistema nervioso es un detector de señales externas en continua actividad con un telón de fondo de otras señales internas y externas que llegan en medio de procesos de pensamiento continuo, ruido sensorial. El problema del ruido sensorial, es decir, el error aleatorio al percibir los estímulos, es que produce efectos variables a lo largo de los ensayos. La solución a este problema se realiza aquí mediante la hipótesis Φ - γ , que afirma que los factores que contribuyen al error son independientes, por lo que las variaciones a través de los ensayos son de tipo normal.

El modelo fechneriano se mantiene vigente y sucesivas aportaciones han expandido los conocimientos psicofísicos explorando modos alternativos de estimación directa, es decir, los sujetos responden de forma directa mediante estimaciones subjetivas proporcionando medidas de semejanza de estímulos⁷, utilizando los métodos que exponemos:

- Estimación de magnitud: frente a un estímulo individual, se pide al sujeto que defina su magnitud de forma numérica.

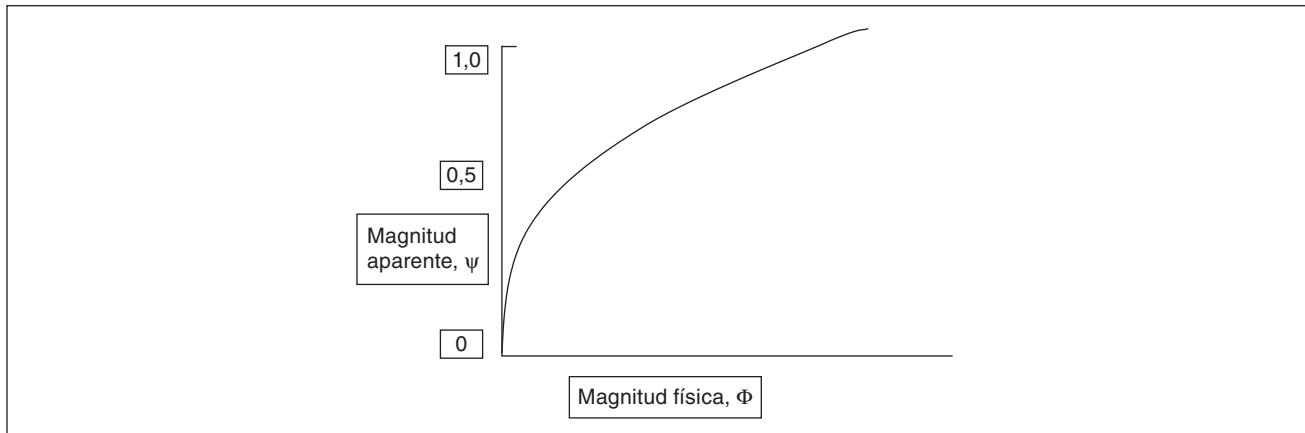


Figura 2. Modelo de Stevens.

- Producción de razón: tras la administración de un estímulo, se demanda al sujeto que produzca otro estímulo que crea que dobla en intensidad al primer estímulo.
- Estimación de razón: al administrar al sujeto un estímulo patrón y un segundo estímulo de comparación, se le pide que valore la razón de sus intensidades.
- Bisección: tras la presentación de dos estímulos, un tercero debe ser ubicado por el sujeto a mitad de la distancia de los dos primeros; este método da lugar a datos interválicos.
- Igualación modal cruzada: un estímulo de una índole debe ser igualado por el sujeto utilizando un estímulo de índole distinta.
- Método de intervalos de apariencia igual: los sujetos deben clasificar los estímulos en categorías de forma que los rangos entre las categorías parezcan iguales.

En S.S. Stevens⁸ la ecuación utilizada (ley de poder) permite que las razones físicas proporcionen las razones subjetivas iguales:

$$\Psi = b \Phi^a$$

donde b es constante escalar y a describe la razón sensorial asociada con la razón física de dos estímulos que difieren a lo largo de la dimensión física que se estudia. Como ejemplo, en el estímulo de la brillantez de las fuentes de luz, podemos estimar que aumenta como la raíz cúbica de la intensidad asociada, de donde se deduce que cuando la razón sea de 8:1 parecerá que la intensidad es el doble del primer estímulo. La gráfica que se obtiene aplicando la ley de poder puede ser como se presenta en la figura 2.

En algún momento del desarrollo de estos conocimientos debe haber una línea que extienda los conocimientos psicofísicos y permita su aplicación a la psicometría. El punto de unión fue determinado por la ley de Fullerton-Cattell, al establecer que las diferencias entre estímulos apreciadas con la misma frecuencia son iguales, a no ser que siempre o nunca sean apreciadas. Es decir, si el 75% de las veces el estímulo A es apreciado como superior a B, y el 75% de veces B es apreciado como superior a C, podemos asumir que la distancia entre A y B, y B y C son iguales⁹. A partir de aquí, la relación «superior a» puede significar valores preferenciales.

Estos sistemas clásicos han supuesto un logro plausible y un método válido para sistematizar la medición de sensaciones basada en juicios sobre los estímulos. Más recientemente, los esfuerzos se han orientado hacia el estudio de la precisión discriminante de los sujetos, en lo que se ha denominado la teoría de la detección de señales¹⁰. De acuerdo con esta teoría, siempre existen otras variables interoceptivas, de forma que durante el proceso de atención sostenida se cometen errores porque los criterios de respuesta se vuelven más conservadores. Para estudiar la probabilidad de juzgar la presencia o ausencia de atributo se utiliza la escala de Thurstone, donde la variabilidad en ausencia de señal (distribución de ruido, r) es la misma que la variabilidad en presencia de señal (distribución señal más ruido, $s + r$).

Las extensiones de esta teoría, como la teoría del reconocimiento general¹², permiten aplicar esta estrategia a los juicios duales. Para este fin se aplican tareas de: 1) estimación (niveles de respuesta > niveles de estímulo); 2) identificación (niveles de respuesta = niveles de estímulo); 3) categorización (niveles

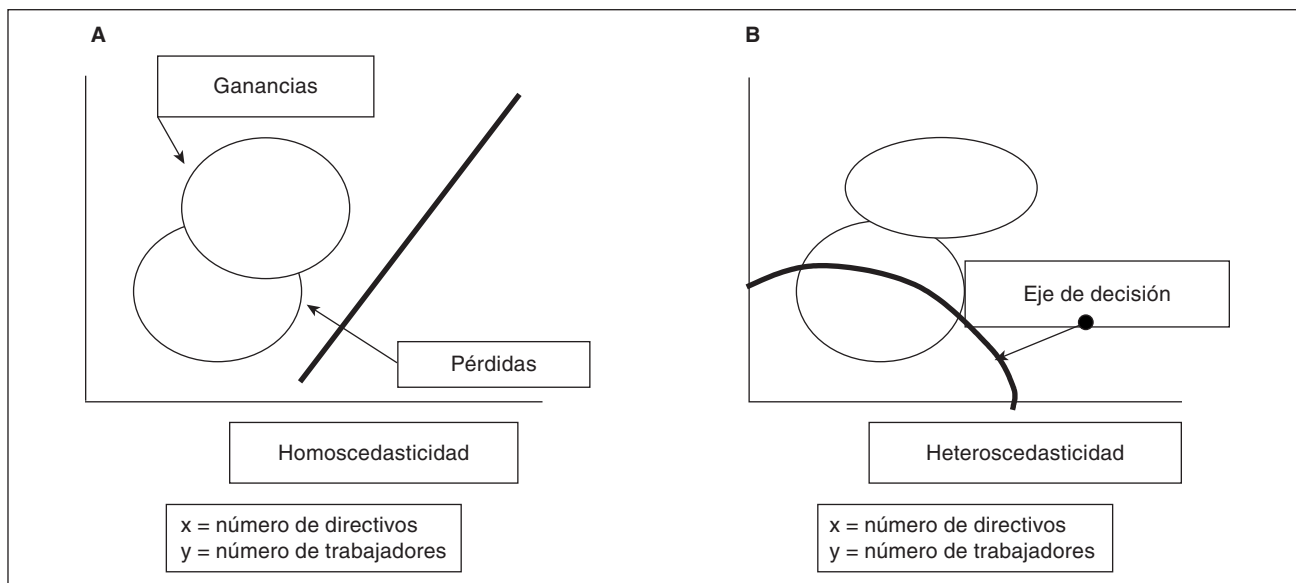


Figura 3. Teoría del reconocimiento general.

de respuesta < niveles de estímulo); 4) filtración (una dimensión es ignorada); 5) condensación (los sujetos juzgan una combinación de atributos), y 6) control (se mantiene una dimensión irrelevante). En la teoría de reconocimiento general se utilizan gráficas de dispersión de estímulos que se agrupan en contornos de probabilidad de tipo geométrico y basado en vectores como la que aparece en la figura 3. La forma de los contornos (esferas o elipses), donde se encuadran los estímulos, permite conocer la independencia o correlación entre las dimensiones. La situación de los contornos en el espacio define la homoscedasticidad (cuando los contornos tienen la misma forma) o heteroscedasticidad (distintas matrices de varianza-covarianza de los estímulos), que indica falta de independencia perceptual. Sobre estos contornos es posible trazar un eje discriminante que predice una función de los estímulos representados o una regla de decisión óptima no lineal en caso de heteroscedasticidad¹². En la figura 3 se presenta un ejemplo de sección discriminante basándose en el número de directivos y trabajadores sobre las ganancias o pérdidas de una empresa.

El modelo de Ashby y Townshend, basado en la representación geométrica de los estímulos, puede producir resultados anómalos en el caso de que aumente la varianza perceptual (un estímulo es más ambiguo o si las dimensiones son valoradas como interdependientes). Estas limitaciones llevaron al desarrollo de modelos no geométricos basados en la identificación de vecindad entre estímulos; el estímulo más próximo es el que tiene la estimación de semejanza media más

cercana a otro estímulo¹³. La forma de representación para este modelo es conocida como árbol aditivo, dado que la geometría restringe el número de puntos que pueden tener un vecino más próximo, la distancia entre estímulos es proporcional a las distancias horizontales en el árbol (las verticales son indiferentes), las localizaciones de los estímulos son denominadas nodos y sus conexiones como trayectos. A raíz de sus estudios se definieron los parámetros: 1) centralidad (C) refleja el grado en que unos estímulos dominan a otros, y 2) reciprocidad (R) describe el grado en que la relación entre estímulos es simétrica. La figura 4 ilustra algunos tipos de árboles aditivos que se utilizan para describir la semejanza.

Estos sistemas modernos de valoración han supuesto la puesta en marcha de métodos estandarizados de valoración del dolor, ayudando a comprender mejor sus bases neurofisiológicas, utilizando estímulos controlados aplicados a paradigmas como umbral del dolor, discriminación, detección y comportamiento¹⁴, mecanismos neurológicos de inhibición del dolor y la eficacia de los distintos tratamientos del dolor¹⁵.

Como conclusión, hemos visto que la tradición fechneriana estudia los juicios frente a estímulos ordinales basándose en la comparación pareada; la relación entre magnitud física del estímulo y la sensación que evocan responde a una función logarítmica. El punto de vista de S.S. Stevens deja a los individuos que emitan juicios sobre intervalos o razones de magnitud entre estímulos; la relación entre estímulo y sensación percibida es una función de

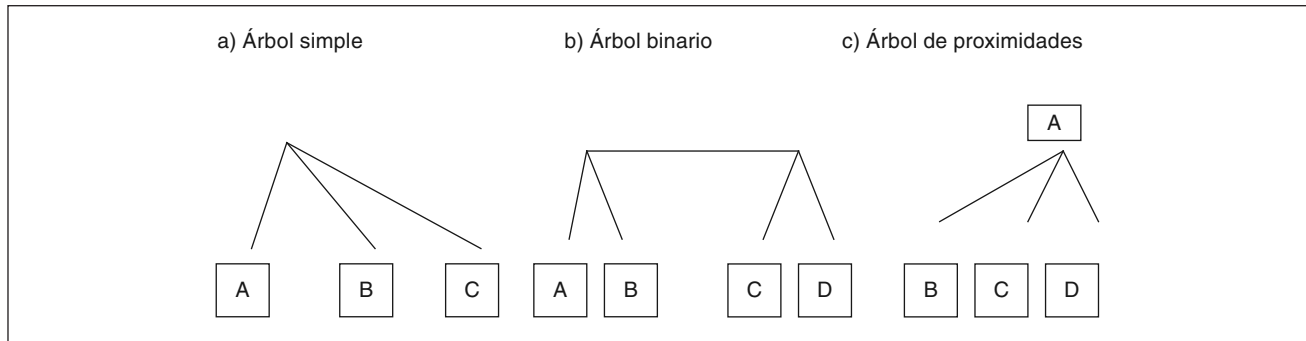


Figura 4. Distintos árboles aditivos.

poder. El vínculo entre psicofísica fechneriana y la psicometría nace de la ley de Fullerton-Cattell, y los modelos más recientes asumen supuestos más realistas, como se observa en la teoría de la respuesta al ítem (IRT) que se enuncia más adelante. Se han citado, además, los modelos para el análisis de datos categóricos y clasificación binaria. Tras la identificación de que algunos sujetos detectan estímulo cuando éste no se ha producido y viceversa, la teoría de la detección de señales se centra en el estudio la capacidad discriminante usando los modelos de Thurstone. En la teoría de reconocimiento general, mediante gráficas de dispersión de estímulos, se estudia la probabilidad de independencia o correlación. En el caso de ambigüedad de los estímulos, los subsiguientes desarrollos se basan en la vecindad entre estímulos, mediante el árbol aditivo para el estudio de la dominancia o interdependencia entre estímulos.

MODELOS DE ANÁLISIS DE LAS PUNTUACIONES

Como se ha podido observar hasta este punto, existen procedimientos muy adecuados para integrar los conocimientos y hallazgos biomédicos. Estos procedimientos utilizan escalas en las que se integran diversos ítems, o ítems presentes en la observación y operaciones matemáticas, de las que derivan las puntuaciones finales de la escala. Una vez obtenida ésta, disponemos de la metodología para determinar el funcionamiento de cada uno de los ítems (análisis de ítems), y de la interpretación normativa de los resultados obtenidos mediante la prueba (estandarización de pruebas). Adicionalmente, se puede analizar la adecuación de las operaciones matemáticas; las matemáticas consisten en una casi ilimitada manipulación de símbolos que, basándose en unas reglas,

producen series deductivas que no tienen por qué representar eventos presentes en el mundo real.

Teoría clásica de la prueba

De acuerdo con la teoría clásica de la prueba, la puntuación es el resultado de un sumatorio, quizás ponderado, de las respuestas obtenidas en cada ítem. Las limitaciones que se han presentado son las siguientes:

- Las puntuaciones no discriminan entre los valores individuales de los apartados que la componen y que evalúan aspectos conceptualmente distintos y potencialmente independientes. A este fenómeno se le conoce como efecto de halo¹⁶.
- No provee de información sobre adecuación de la operación utilizada para obtener las puntuaciones (v. g. sumatorio), en un contexto en que la combinación de ítems con distinto peso puede dar lugar a puntuaciones finales idénticas usando distintos ítems de la escala.
- Al depender las puntuaciones de una muestra las medidas, como los coeficientes de fiabilidad y de validez, también dependen de la muestra.
- Las medidas creadas por autoevaluación no se correlacionan exactamente con otros procedimientos como observación de comportamiento o exámenes físicos, por lo que no es de esperar que en una segunda administración podamos obtener el margen de error de la prueba.

Uno de los desarrollos psicométricos más interesantes y modernos deriva de la constatación de que el peso de los ítems que componen la escala puede no ser homogéneo; algunos de los ítems pueden contener un sesgo o error inherente superior a otros y la puntuación total obtenida puede no ser apropiada para computar los resultados de algunos pacientes.

Teoría de la respuesta al ítem

La teoría de respuesta al ítem (*item response theory*) utiliza el patrón de respuesta para responder a la pregunta sobre si las operaciones matemáticas para obtener los valores de escala son justificables o apropiadas. Dada una relación entre un atributo a medir y la posibilidad de una respuesta determinada, se obtiene una línea de trazo para cada ítem; la localización (umbral) determina la dificultad de esa respuesta, mientras que su inclinación determina la capacidad de ese ítem para discriminar. Las diferencias respecto a su predecesora son:

- Los ítems de una escala se evalúan de acuerdo con su sesgo, nivel de dificultad, capacidad discriminante o relación con otros conceptualmente similares recogidos por métodos diferentes. Así, la teoría de la respuesta al ítem no se basa en valores sumatorios o ponderados como la teoría clásica sino en patrones de respuesta.
- Los métodos de puntuación dan lugar a medidas independientes de la escala que permiten que puntuaciones de distintas escalas puedan contrastarse para obtener mejores interpretaciones, utilizando una familia de algoritmos matemáticos que, por definición, son independientes de la escala y de desviaciones.
- Los parámetros no dependen de la muestra. Para ello se usan bancos de ítems ya calibrados, seleccionando los ítems óptimos para medir construcciones específicas en situaciones específicas.
- No es necesario repetir las mediciones para analizar su precisión, debido a que los parámetros personales no son definidos como valores repetibles, sino como parámetros-modelo. En la teoría de la respuesta, los análisis clásicos de fiabilidad, como correlaciones test-retest, han sido sustituidos por análisis estadísticos de las cualidades de las estimaciones de los parámetros modelo.
- Analizan los modelos matemáticos utilizados para determinar si los datos obtenidos son apropiados para un modelo específico e indica qué operaciones pueden ser utilizadas de forma legítima para esos datos.

Para conseguir estos propósitos, la teoría de la respuesta al ítem relaciona en una línea de trazo, donde se encuentra representada la habilidad que se pretende medir (θ) con la probabilidad de una respuesta designada. Para estimar θ asume un patrón de respuestas del sujeto. Una vez controlada la respuesta θ , los ítems se vuelven independientes, lo que se

conoce como independencia condicional. Los valores de θ pueden caer dentro de un continuo (teoría de rasgo latente) o formar categorías más discretas (teoría de clase latente). Como ejemplo de rasgo latente podemos considerar un supuesto en que cinco partidos políticos representan distintas fuerzas sociales en dos elecciones sucesivas; su elección puede estar inferida por clases latentes (podrían ser pertenencia a derechas o izquierdas o una cuestión social importante) o rasgos latentes (las políticas aplicadas sobre educación, sanidad o bienestar valoradas como un continuo). Las informaciones sobre el rasgo seccionan toda asociación. Si conocemos las preferencias en las primeras elecciones podemos conocer algo sobre la clase de los votantes; una vez conocida esa clase las elecciones se vuelven independientes.

La teoría de respuesta al ítem puede aplicarse en algesiometría siempre que el instrumento utilizado permita dar una sola respuesta a cada ítem, utilizando una serie de posibles respuestas que pueden ordenarse de forma ascendente.

Modelos monoparamétricos

El modelo más sencillo de la teoría de rasgo latente logístico es el descrito por el matemático danés G. Rasch, también llamado modelo monoparamétrico (asume que los ítems varían sólo en su dificultad pero son igualmente discriminantes), aunque existen modelos biparamétricos que analizan la capacidad discriminante del ítem. Este modelo estima un atributo dado el cual aquellos sujetos que están por debajo de un determinado valor umbral (para ese atributo) tendrán poca probabilidad de un tipo de respuesta, y los que están por encima de él casi siempre elegirán una determinada respuesta. Tanto este modelo como el biparamétrico ignoran la adivinación, y el modelo triparamétrico incorpora la tasa de falsos positivos, es decir, la capacidad de adivinar sin poseer el atributo que se está midiendo. Este modelo se ha extendido a lo que se conoce con el nombre de modelo de crédito parcial (PCT, *partial credit testing*)¹⁷. La noción PCT es importante dado que progresiones de estímulos crecientes (como pueden ser molesto-moderado-insoportable) suponen que la elección de uno de ellos implica que los que le preceden son sobrepasados, mientras que los que le suceden no lo son. El PCT computa una probabilidad condicional, como una función de la persona, el parámetro y el lugar que ocupa en la progresión. El modelo de Rasch permite al instrumento de medición funcionar independientemente del objeto a medir, a la vez que da una estimación de los términos necesarios para establecer las calibraciones.

Utilizando el modelo de Rasch, las personas y los parámetros pueden tratarse matemáticamente por separado. Los cálculos de probabilidad se resuelven según el procedimiento de estimación de máxima verosimilitud. En el análisis de Rasch, estos análisis están dirigidos a fraccionar cada respuesta del paciente en el instrumento sin tener en cuenta quién lo responde, y el componente como estimación personal sin tener en cuenta el parámetro utilizado. Se ha demostrado un aumento de la sensibilidad de diversos instrumentos cuando se utiliza este sistema de puntuación¹⁸.

El cuestionario de dolor de McGill (MPQ, *McGill Pain Questionnaire*) cumple la legitimidad estructural para ser analizado como un PCT. Al contestarlo, sólo se permite elegir un descriptor dentro de una subclase en donde los descriptores están ordenados en un orden ascendente. Este método puede ayudar a superar los problemas que derivan de la transformación de las cualidades presentes en una puntuación final que refleja el abstracto, de forma que iguales puntuaciones pueden reflejar cuadros distintos, y, por otra parte, el hecho de que los valores de los ítems individuales dentro de una subclase no son aritméticamente equivalentes a los valores de otros descriptores en otras subclases, es decir, la diferencia entre cualquier par de descriptores adyacentes (o entre no respuesta y el primer descriptor de la progresión). La utilización de la teoría de la respuesta al ítem para obtener puntuaciones utilizando el MPQ mostró resultados prometedores en diversos estudios^{19,20}.

Modelos biparamétrico y triparamétrico

Si en el modelo monoparamétrico se asume que los ítems varían sólo en su dificultad pero son igualmente discriminantes, el modelo biparamétrico (2PL) analiza además la capacidad discriminante del ítem. Al igual que el modelo monoparamétrico, ignora la capacidad de adivinación del sujeto, es decir, la probabilidad de que alguien que posee el atributo a medir en poca cuantía responda al ítem de forma correcta. Para superar esta limitación existe el modelo triparamétrico (3PL), que incorpora la tasa de falsos positivos, es decir, la capacidad de adivinar, y es apropiado para las pruebas de elección múltiple y de verdadero-falso. También se han desarrollado modelos no paramétricos²¹.

PARADIGMAS PARA LA VALORACIÓN DE LOS MÉTODOS PSICOMÉTRICOS

Cualquier método diagnóstico está sometido a un grado de inexactitud propia de la interferencia que

provoca el observador sobre el hecho que se observa. Todo acto de observación contiene variables que van desde de la naturaleza de la muestra, el procedimiento utilizado y la interpretación de los datos obtenidos.

Las puntuaciones de una medición se entienden como una combinación de verdad y algún error de medición o de tipo aleatorio. La interpretación de los resultados extraídos por métodos psicométricos depende de la comprensión de la estructura del propio método²².

Para que los resultados obtenidos por valoración psicométrica sean considerados útiles en situaciones prácticas, su diseño debe satisfacer ciertas condiciones básicas, como fiabilidad y validez. De forma genérica, un instrumento será fiable si es capaz de producir estimaciones repetibles, y será válido si representa el área de interés en el estudio.

Fiabilidad

Este concepto determina la proporción de personas en estudio que son correctamente clasificadas por el instrumento de evaluación. Para ello las subsecuentes mediciones deben dar lugar a mediciones repetibles. Una medición fiable tiene la propiedad de proveer de resultados consistentes bajo circunstancias variantes. La replicabilidad, reproducibilidad y consistencia pueden considerarse como sinónimos. Conviene diferenciar este concepto del de precisión: una báscula puede dar lugar a mediciones no muy precisas, pero aun así ser fiable si el grado de desviación es constante, o, contrariamente, producir mediciones muy precisas pero no fiables en cuanto a que se afecten por otras variables (v. g. altitud sobre el mar).

Para conocer el grado de fiabilidad se utilizan distintos tipos de análisis, y normalmente se empleará más de uno. Los más empleados son:

- Fiabilidad en test-retest (estabilidad): el objeto de este análisis es determinar si aplicaciones repetidas del instrumento en el mismo sujeto o grupo de sujetos tiende a cambiar los resultados. Este análisis cobra importancia cuando se trata de realizar mediciones con fines predictivos.
- Fiabilidad entre evaluadores (equivalencia): este análisis se realiza comparando los resultados obtenidos en un sujeto o grupo de sujetos por dos o más evaluadores. La técnica más utilizada es la correlación intraclases o índice de concordancia de Kendall.
- Consistencia interna: estudia la homogeneidad de los apartados que forman el instrumento psico-

métrico. Asumiendo que una medición en la que sus apartados tienen una alta correlación produce resultados más consistentes, la razón de introducir más de una pregunta sobre un aspecto concreto evita la posibilidad de que respuestas de tipo idiosincrásico produzcan desviaciones excesivas en las puntuaciones. Las fórmulas más utilizadas en este tipo de análisis son:

- La técnica de división en dos mitades: este método consiste en dividir el cuestionario en dos partes y observar la correlación entre los ítems presentes en cada mitad. Este sistema resulta adecuado para escalas que utilizan varios ítems donde todos y cada uno de ellos miden uno mismo constructo. La razón por la que se generan varios ítems en este tipo de escalas es la de reducir la posibilidad de respuestas idiosincrásicas que puedan afectar a la medición realizada, produciendo una mayor precisión. En el caso de MPQ este enfoque resulta inadecuado, ya que los ítems presentes (subclases) miden distintos componentes posibles del dolor, pero no siempre presentes.
- El coeficiente α de Cronbach: es un valor derivado de considerar todas las correlaciones entre los valores de los ítems presentes e informa sobre la puntuación media de éstos. Un valor de 1 indica concordancia perfecta entre los valores de los ítems. Cuando el valor es muy bajo, se puede inferir que los distintos ítems no están midiendo el mismo atributo.

Para realizar estos análisis se utiliza básicamente el coeficiente de correlación (r), obteniendo mediciones al mismo sujeto en dos entrevistas separadas en el tiempo (estabilidad), o en el mismo sujeto por parte de varios entrevistadores (equivalencia). En la α de Cronbach es una fórmula específica que no es el coeficiente de correlación de Pearson.

Validez

Este parámetro expresa el grado en que una medición se aproxima a la cantidad o calidad real de la entidad en estudio. Estudia la propiedad del cuestionario como expresión de la entidad en estudio. Las cualidades psicométricas más utilizadas para determinar la validez de una medición son²³:

- Validez de contenido: establece si la medición está formada por componentes representativos de la entidad a medir, es decir, si son relevantes. Uno de los aspectos a los que se refiere es la validez

aparente, donde un ítem aparece *a priori* relevante para ser incluido en los ítems. La decisión sobre su inclusión está basada en la experiencia personal o la de un asesor y en una construcción teórica. Otro aspecto a tener en cuenta, para finalizar, es la comprensión de los ítems presentes.

Entre los tipos de validez basada en la relación con otras variables se encuentran las siguientes:

- Validez convergente: este término estudia la propiedad de la escala de producir mediciones de forma armónica con otras escalas preexistentes.
- Validez divergente o discriminante: las relaciones entre una prueba y las medidas que supuestamente evalúan constructos distintos proporcionan información sobre la validez discriminante.

Pueden utilizarse análisis de precisión para indicar la capacidad de un determinado proceder diagnóstico para discriminar entre dos subclases de sujetos, o conocer la proporción de sujetos correctamente clasificados por una determinada prueba (siempre que no sea dicotómica). Este análisis requiere trazar una curva de operación característica del receptor (ROC), donde se relacionan las tasas de verdaderos aciertos (sensibilidad en eje de ordenadas) y de falsos positivos (1 especificidad en eje de abscisas), cada punto representa un par sensibilidad (1 especificidad) correspondiente a un criterio de decisión, y la curva examina todos los umbrales de decisión para todos los resultados observados (Fig. 5).

Una prueba diagnóstica de gran precisión estaría situada en la parte superior izquierda, donde hay más aciertos (gran sensibilidad) y hay menos falsos positivos (gran especificidad). Normalmente se expresan en forma de transformaciones z ; aquí las transformaciones z derivan de índices de correlación $r^{1,11}$ (Fig. 5).

- Validez de criterio son la concurrente y la predictiva.
 - Validez concurrente: determina el grado en que la nueva medición se correlaciona con otros procedimientos ya aceptados en el mismo periodo de tiempo.
 - Capacidad predictiva: estudia la capacidad de una prueba para determinar si los sujetos que producen determinados perfiles en la prueba desarrollarán una condición en el futuro.
- Validez de constructo: valora la estructura de varias mediciones relacionadas entre sí por la entidad que se desea conocer. Es una forma multivariante de validez concurrente que se aproxima

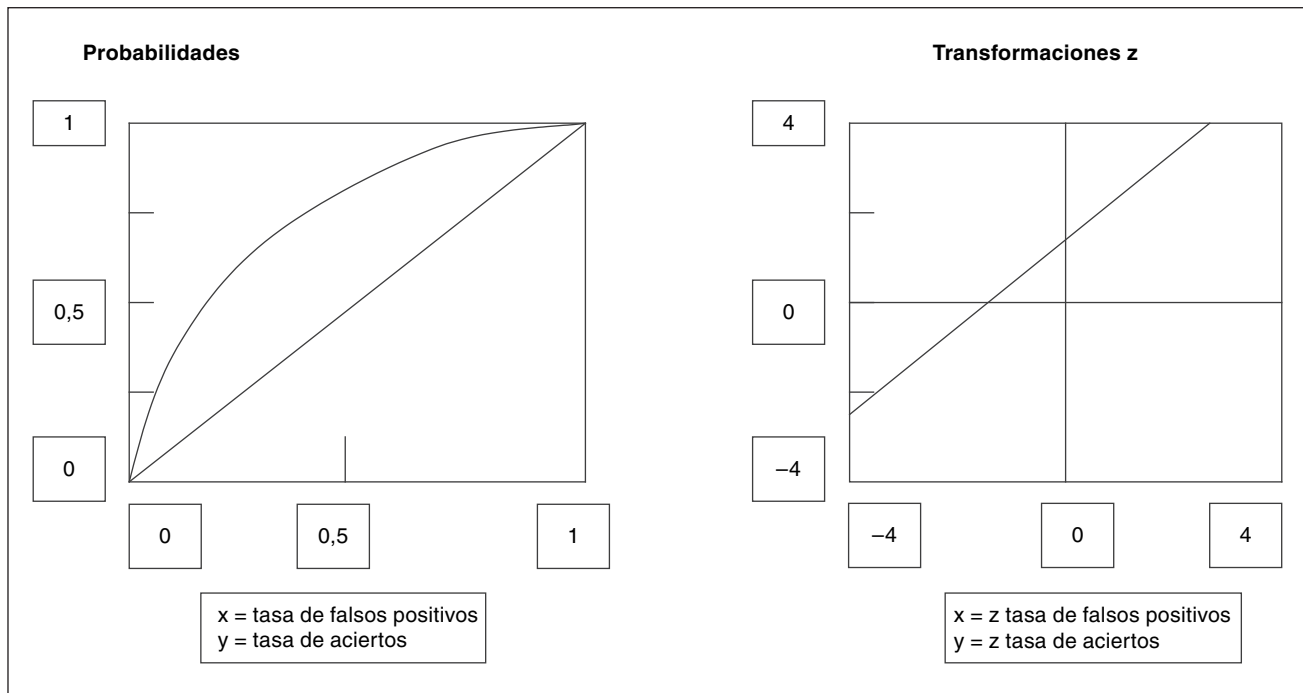


Figura 5. Curva de operación característica del receptor (ROC).

en mayor grado al mundo multidimensional objeto de nuestras investigaciones. Considerando el aumento de variables que se dan encuentro en este proceso, se precisa técnicas multivariantes específicas que incluyan estimaciones del error presente en las mediciones. Para ello, las escalas se pueden someter a un análisis factorial. Éste es un método estadístico que analiza todas las posibles intercorrelaciones de los ítems en una prueba para determinar si esos ítems o apartados caen dentro de grupos, o factores, que pueden emerger como componentes de la entidad que se pretende medir. Recientemente se han incluido nuevas normas para el estudio de la validez²⁴:

- Redundancia: dos ítems pueden dar una información idéntica sobre la construcción teórica.
- Correlación: dos ítems pueden estar altamente correlacionados.
- Irrelevancia: un ítem que aporta muy poca información sobre la construcción que se pretende explorar, puede considerarse irrelevante.

Así, el análisis factorial está encaminado a tres grandes grupos de cuestiones: 1) reducir la cantidad de ítems necesarios para definir el sistema que se pretende medir, sin perder precisión; 2) hallar las distintas cualidades del sistema que se está midiendo, identificando grupos consistentes de ítems que miden cualidades independientes dentro del sistema, y 3) explorar a qué grupo (factor) se pueden adscribir cada uno de los ítems.

ANÁLISIS FACTORIAL

El análisis factorial puede ser de gran utilidad en el estudio de la validez de constructo.

Una vez diseñado un cuestionario, los autores pueden estar interesados en reconocer las cualidades representadas en los ítems y perfeccionar el cuestionario retirando algunos de ellos, ya que los ítems que aparecen pueden hallarse en tres situaciones que los hacen prescindibles:

La técnica empleada por estos sistemas de análisis consiste en computar una matriz de correlación para todas las variables que forman la muestra de datos. Dentro de esta matriz de correlación, se crearán unas nuevas variables que definen esta matriz; estas variables son independientes entre ellas (factores). Cada uno de estos factores (*eigenvectors*) lleva asociado un valor propio (*eigenvalue*), que representa la cantidad de varianza representada por esta variable en la muestra. Así, se obtiene una lista de las variables originales, y al lado de cada una de ellas figura un

valor propio y la cantidad de varianza que depende de dicha variable. En una segunda operación se obtiene la correlación de cada una de las variables con los factores obtenidos. Para mejorar el contenido de un cuestionario es habitual:

- Ignorar las correlaciones pequeñas (por convención, inferiores a 0,3), ya que suponen sólo el 9% de la varianza observada.
- Ignorar las variables que obtienen correlaciones superiores a 0,3 en varios factores. Seguramente no se corresponden de forma unívoca con ninguno de ellos.

Después de la obtención de una primera solución factorial se puede proceder a la rotación de los factores iniciales. Este posicionamiento distinto de los ejes abundará en la diferencia observada en la correlación de cada variable con los factores. Un sistema muy utilizado es la rotación Varimax con el criterio de Kaiser, para seleccionar el número de factores significativos. Existen más sistemas de rotación, algunos ortogonales y otros oblicuos.

Finalmente, conviene recordar que los factores identificados por este sistema describen los datos introducidos en la muestra y no son necesariamente generalizables, si bien supone un método plausible para identificar los componentes del sistema de medición en una determinada muestra.

ESCALAMIENTO MULTIDIMENSIONAL

Además del análisis factorial existen otros sistemas para reconocer las semejanzas entre los ítems y establecer jerarquías de grupos de ítems dentro del instrumento.

Algunos objetos de estudio son inaccesibles a la observación y se requiere alguna técnica que permita encontrar las dimensiones subyacentes de estos fenómenos y generar teorías basándose en los juicios que se pueden extraer de los pacientes. Dentro de los métodos que sirven al escalamiento multidimensional encontramos el escalamiento de diferencias individuales (INDSCAL, *individual differences scaling*) y el mapeo de preferencias (PREFMAP, *preference mapping*).

Espacio del grupo de estímulos

Supongamos la existencia de un mundo en el que están representados todos los aspectos relacionados

con la experiencia dolorosa, el cual podríamos llamarlo «dolor global», pero nosotros desconocemos las dimensiones de este mundo. Conocemos determinados acontecimientos que ocurren en este mundo y a cada uno de ellos lo llamaremos estímulo objetivo. Pueden ser físicos o descriptores verbales (palabras que en este caso describen algunas cualidades del dolor). Podemos presentar pares de estímulos a unos sujetos y preguntarles hasta qué punto creen que esos estímulos se parecen. Ésta es una de las ideas clave para el escalamiento multidimensional; la similitud de juicios emitidos representa la distancia psicológica. Cada uno de los estímulos objetivos representa un punto del espacio multidimensional abstracto y se situará basándose en su similitud.

El procedimiento INDSCAL determina las distancias entre pares de estímulos objetivos (espacio de distribución de estímulos) y es capaz de situar a un individuo basándose en la relevancia de cada una de las dimensiones para un individuo (espacio individual)²⁵. Se obtiene un gráfico de dispersión de puntos (espacio del grupo de estímulos) cuyas coordenadas revelan la relación entre estímulos en un espacio de «r» dimensiones; el orden de los estímulos permite hacer teorías sobre la cantidad de dimensiones subyacentes y la importancia o prominencia de esa dimensión para un individuo determinado²⁶.

Buscando un ejemplo geográfico, medir distancias en un mapa resulta una tarea simple, pero generar un mapa basándose en distancias entre puntos resulta algo más complicado. Esto es lo que realiza INDSCAL; toma matrices de similitudes subjetivas que contienen cierto grado de error y genera distancias sobre un mapa. Supongamos que a un grupo de sujetos se le pide que puntúe la similitud entre seis ciudades (objetos); cada sujeto generará una matriz de 15 celdas. Si sus respuestas se basan solamente en la población de las ciudades, emergerá una sola dimensión; si además sus juicios se basan en las distancias, la población y la altura sobre el nivel del mar, aparecerán cuatro dimensiones, de forma que el número de dimensiones viene determinado por los factores representados en la matriz. No se requiere un conocimiento *a priori* de la dimensionalidad; al contrario, la configuración espacial es una evidencia de las dimensiones que son relevantes para los sujetos. En suma, dado un juicio de similitudes entre pares de estímulos objetivos, se obtienen distancias entre ellos: a mayor grado de disimilitud entre objetos, mayor es la distancia en el espacio abstracto que se resuelve obteniendo una matriz. Los datos obtenidos por este análisis son interpretados por simple

inspección de la configuración establecida en los puntos o sometiendo la matriz obtenida a análisis de *clusters*²⁷.

En el terreno del dolor se hallaron tres dimensiones en el espacio del grupo de estímulos que en rasgos generales concuerdan con las categorías de dolor presentes en el MPQ, si bien se ha sugerido que debería aumentarse el número de descriptores en las categorías afectiva y evaluativa dado que éstas resultaron más prominentes que la categoría somatosensorial^{26,28}. En modelos de dolor experimental se han identificado tres dimensiones en relación con las estrategias utilizadas para afrontar el dolor inducido por método criopresor. Éstas eran el reconocimiento de la sensación, la relevancia de actitud y un componente cognitivocomportamental²⁹.

El propio método de escalamiento de diferencias individuales ha permitido reconocer que los juicios subjetivos utilizados para hallar las distancias entre pares de estímulos pueden estar condicionados por la actitud y percepción de los individuos seleccionados en el estudio³⁰. Otras limitaciones de este sistema derivan de asumir que todos los puntos (estímulos) representados en el espacio son equivalentes e igualmente variables.

Mapeo de preferencias

En algunos casos se requiere una interpretación más profunda de la información proveniente de INDSCAL como, por ejemplo, cuando las dimensiones son ambiguas o se quiere conocer el significado de las dimensiones creadas. En esas situaciones el mapeo de preferencias (PREFMAP) computa un vector basado en las propiedades y/o atributos psicológicos de cada objeto (estímulo) usado en el análisis INDSCAL.

Siguiendo con el ejemplo geográfico, se puede pedir al sujeto que puntúe de 0-10 todas las ciudades (objeto) previamente presentadas en el análisis de INDSCAL contra escalas de propiedad bipolar que consisten en los atributos de posible interés (v. g. nivel del mar-elevación, calor-frío, pequeña-grande). De esta forma, cada sujeto examina todas las dimensiones presentes en el análisis INDSCAL previo y las propiedades de la escala.

BIBLIOGRAFÍA

1. Nunnally CJ, Bernstein IJ. Teoría psicométrica. México: McGraw Hill, 1995; p. 769-91.
2. Berriman AE. Historical metrology. Londres: Dent; 1953.
3. Barrow JD. The constants of nature. Londres: Vintage; 2002.
4. Bridgman PW. The logic of modern physics. Nueva York: McMillan; 1928.
5. Maturana H, Varela F. El árbol del conocimiento. Santiago de Chile: Editorial Universitaria; 1984.
6. Bock RD, Jones LV. The measurement and prediction of judgement and choice. San Francisco (CA): Holden Day; 1968.
7. Stevens SS. Mathematics, measurement and psychophysics. En: Stevens SS, ed. Handbook of experimental psychology. Nueva York: Wiley; 1951.
8. Stevens SS. On the theory of scales of measurement. Science 1946; 103:677-80.
9. Fullerton GS, Cattel JM. On the perception of small differences. Filadelfia (PA): University of Pennsylvania; 1929.
10. Tanner WP, Sweets JA. A decision-making theory of visual detection. Psychol Rev 1954;61:401-9.
11. Lusted LB. Decision-making studies in patient management. N Engl J Med 1971;284(8):416-24.
12. Ashby FG, Townshend JT. Varieties of perceptual independence. Psychol Rev 1986;93:154-79.
13. Tverski A, Hutchinson JW. Nearest neighbour analysis of psychological spaces. Psychol Rev 1987;93:3-22.
14. Price DD, Harkins SW. Psychological approaches to pain measurement and assessment. En: Turk D, Melzack R, eds. Handbook of pain assessment. Toronto: Guilford Press; 1992. p. 111-34.
15. Gracely RH. Psychophysical assessment of human pain. En: Bonica JJ, Lieberskind JC, Albe Fessard DG, eds. Advances in pain research and therapy. Vol 3. Nueva York: Raven Press; 1979.
16. Wells FL. A statistical study of literary merit. Arch Psychol 1907;1:7.
17. Wright BD, Masters GN. Rating scale analysis: Rasch measurement. Chicago (IL): Mesa Press; 1982.
18. Norquist JM, Fitzpatrick R, Dawson J, Jenkinson C. Comparing alternative Rasch-based methods vs. raw scores in measuring change in health. Med Care 2004;42 Suppl 1:25-36.
19. Kalinowski AG. Measuring clinical pain. J Psychopathol Behav Assess 1981;7:329-49.
20. McArthur DL, Cohen MJ, Schandlers SL. A philosophy for measuring of pain. En: Chapman C, Loeser JD, eds. Advances in pain research and therapy. Nueva York: Raven Press; 1989. p. 37-50.
21. Mokken RJ, Lewis C. A non-parametric approach to the analysis of dichotomous responses. Appl Psychol Measurement 1982;4:1:341-7.
22. Mann AH. The clinical relevance of psychometric testing. Eur J Clin Pharmacol 1985;28:32-4.
23. Dempsey PA, Dempsey AD. Nursing research: text and notebook. Boston (MA): Little & Brown; 1996.
24. National Council on Measurement in Education. Standards for Educational and Psychological Testing. Washington DC: AERA, 1999.
25. Carroll JD, Chang JJ. Analysis of individual differences in multidimensional scaling, via an N-way generalization of an "Eckart Young" decomposition. Psychometrika 1977;35:283-319.
26. Crawford-Clark W, Ferrer-Brechner T, Janll ML, Carrol J, Yang J. The dimensions of pain: a multidimensional scaling comparison of cancer patients and healthy volunteers. Pain 1989;37:23-32.
27. Wexler KN, Romney AK. Individual variations in cognitive structures. Multidimensional scaling. Vol 2. Nueva York: Seminar Press; 1972.
28. Carasso RL, Yehuda S, Mostofsky DI. Multidimensional scaling of pain experiences. Bull Br Psychol Soc 1984;22:349-50.
29. Wack JT, Turk DC. Latent structure of strategies used to cope with nociceptive stimulation. Health Psychol 1984;3:27-43.
30. Janal MN, Clark WC, Carroll JD. Multidimensional scaling of painful electrocutaneous stimulation: INDSCAL dimensions, signal detection theory indices, and the McGill Pain Questionnaire. Somatosensory Motor Res 1993;10:31-9.