

REVISIÓN DE LOS MÉTODOS DE ESCALAMIENTO PSICOLÓGICO

MARÍA FÁTIMA DOS SANTOS

Escuela de Psicología-Universidad Central de Venezuela
fatimadossantos@rdsinmarket.com.ve

Resumen

El proceso de medición de magnitudes es fundamental en cualquier ciencia y en psicología se enfrenta con la dificultad que deriva observar directamente las variables medidas, lo cual ha llevado a mucha discusión en torno a la construcción de instrumentos. En este artículo repasamos los procedimientos más conocidos, partiendo del escalamiento psicofísico (que busca la cuantificación de la percepción de estímulos) hasta el escalamiento psicológico (que intenta dimensionar procesos internos que no responden a estímulos directos). Revisamos las conocidas escalas Thurstone, Likert, Guttman, Bogardus, Combs y Osgood, debatiendo las razones por las cuales la discusión sobre escalamiento ha venido perdiendo espacio en las publicaciones científicas, así como explorando posibles líneas de desarrollo en el área.

Palabras clave: teoría de la medición, psicometría, escalas psicológicas, escalas Likert.

Recibido: 15 de marzo de 2014
Aceptado: 16 de septiembre de 2014
Publicado: 30 de diciembre de 2014



REVIEW OF THE METHODS OF PSYCHOLOGICAL SCALING

MARÍA FÁTIMA DOS SANTOS

Escuela de Psicología-Universidad Central de Venezuela

fatimadossantos@rdsinmarket.com.ve

Abstract

The process of magnitudes measurement is fundamental in any science, and in psychology, it faces the impossibility to directly observe the measured variables, which has led to a large discussion around the construction of psychological instruments. In this article we review the most widely known methods, from psychophysical scaling (which seeks to quantify the perception of stimuli), to psychological scaling (trying to measure internal processes that do not respond to direct stimuli). We review the known Thurstone, Likert, Guttman, Bogardus, Combs and Osgood scales, and we discuss the reasons why the scaling procedures discussion has progressively lost ground within the scientific literature. Finally, we explore possible lines of development within the area.

Key words: Likert scale, measurement theory, psychological scales, psychometrics

Received: Mar 15, 2014

Accepted: Sept 16, 2014

Published: Dec 30, 2014

1. GENERALIDADES DEL CONTEO Y LA MEDICIÓN

El conteo y la medición se han desarrollado espontáneamente y con diversos grados de sofisticación en todas las culturas, pues son funciones básicas que brindan sustento a intercambios con lo natural y lo social, y a cierto tipo de comunicaciones entre humanos. Un cazador necesita comunicar a qué distancia está una presa, un comerciante necesita indicar qué cantidad de cabezas de ganado piensa vender, es necesario contabilizar la edad, el peso, etc. (Wright, s.f.). Con el desarrollo de las sociedades, el conteo y la medición se hacen indispensables para la supervivencia en entornos culturales cada vez más complejos, por eso los árabes desarrollaron los números llamados arábigos, los mayas conceptualizaron el cero y los chinos inventaron el ábaco.

La medición es generalmente considerada como “proceso por el cual se asignan números o símbolos a atributos de entidades del mundo real de tal forma que los describa de acuerdo con reglas claramente definidas” (Fenton y Pfleeger, 1997) o, lo que es más o menos lo mismo, la atribución de números a las propiedades de un objeto (Boring, 2004), y es un elemento fundamental del quehacer científico general cualquiera sea la disciplina.

La medición supone la existencia de normas de correspondencia entre lo medido (objetos o sujetos) y el valor que se le asigna. Si consideramos que los elementos de la realidad son miembros del conjunto X y sus correspondientes medidas forman parte del conjunto Y , tenemos una función sobreyectiva, donde a cada elemento de X le corresponde un solo elemento de Y , mientras que Y puede ser imagen de varios elementos de X , como se observa a continuación:

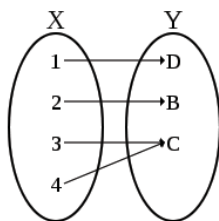


Figura 1. Esquema de la relación entre sujetos y sus medidas

Supongamos que X trata de diferentes sujetos humanos e Y es un conjunto formado por las medidas que se les asigna en la variable estatura. Vemos que cada sujeto tiene su propia medida de estatura y que varios sujetos pueden coincidir teniendo la misma medida.

Por supuesto, existe infinidad de otras dimensiones a las cuales podríamos asignar valores, en las cuales podríamos realizar mediciones: peso, edad, circunferencia craneoencefálica, masa muscular, inteligencia, motivación al aprendizaje, y tal vez, incluso, capacidad de amar o espiritualidad. El hecho de que determinada dimensión pase a ser considerada una variable depende en mucho de la capacidad que tengamos para medirla. Por ejemplo, la espiritualidad será una variable siempre que estemos en capacidad de asignar valores que representen el modo en que esta dimensión ocurre en el mundo real.

Los números asignados durante la medición pueden significar dos cosas muy diferentes:

- a. *El número como etiqueta* permite la clasificación de objetos en categorías. Por razones pragmáticas, las categorías llevan números en lugar de palabras. Por ejemplo, en lugar de clasificar a los humanos en altos, medianos y bajos, podríamos asignarles los números 1, 2 y 3, que significan, respectivamente, altos, medianos y bajos. Pero también podríamos, respectivamente, asignarle los valores 128, 0 y 5497, si nos provoca, porque en este contexto un número mayor no significa mayor cantidad de la cualidad medida. A la asignación de números bajo este enfoque podría llamársele *clasificación*.
- b. *El número como valor*, por el contrario, tiene un significado por sí mismo e indica la magnitud de un determinado sujeto en determinada variable (Boring, 2004). A la asignación de números bajo este enfoque podría llamársele *cuantificación*.

Tanto en un caso como en el otro se aplica por igual el término “medición” como definición general del proceso. Por ejemplo, la asignación del valor “masculino” o “femenino” al género de un grupo de personas tal vez no nos parezca un proceso de medición, porque se trata simplemente de atribución de códigos, pero sí implica medir. La diferencia resulta más clara cuando se revisan los niveles de medición de las variables. Aunque una revisión profunda de ese tema escapa a los objetivos de este texto, tal vez sea conveniente recordar la clasificación más clásica.

Nivel nominal: Consiste en una clasificación de elementos en categorías bien delimitadas (no hay solapamiento entre ellas y los criterios de clasificación son claros). Si se utilizan números, estos sirven como etiquetas y no cuantifican ninguna dimensión. Por ejemplo, al medir la variable “género”, si asignamos el valor “1” a las mujeres y el “2” a los hombres, no quiere esto decir que los hombres son el doble de algo que las mujeres.

Nivel ordinal: Organiza los elementos en un gradiente comparativo, donde sabemos que los que tienen puntuaciones más altas poseen más de determinada cualidad que otros. Sin embargo, la distancia entre los sujetos no es la misma. Un buen ejemplo es el orden de llegada en una carrera: el puesto 1 llegó antes que el 2, y este llegó antes que el 3, pero no podemos saber cuál es la diferencia de tiempo entre estos tres lugares.

Nivel de intervalo: Cumple los requisitos del nivel ordinal, pero además se garantiza que la distancia entre una posición y la siguiente es siempre la misma. Por ejemplo, si dividimos el año en semanas sabemos que la semana 2 ocurre antes que la 6, y también sabemos que entre ellas hay la misma distancia que entre las semanas 20 y 24. Esta escala puede (o no) incluir un valor de 0, pero es un cero arbitrario y no indica la ausencia de cualidad medida. El ejemplo clásico es el termómetro, donde el 0 no expresa una ausencia absoluta de temperatura.

Nivel de razón: Recibe este nombre porque en él es lícito hacer operaciones matemáticas complejas, como la división (o cociente, o razón). Para ello, es necesario que la escala posea un 0 absoluto, es decir, que refleje la ausencia total del valor medido. Por ejemplo, si medimos la distancia entre dos puntos en un plano, esta será de 0 si los puntos están colocados uno encima del otro.

Como vemos, las propiedades de las escalas son aditivas: la escala ordinal clasifica (como hace la nominal) pero además ordena. La escala de razón ordena (como la ordinal), tiene intervalos iguales (como la de intervalo) pero además posee un 0 absoluto.

Sea cual sea el nivel de medición logrado, cuando utilizamos los números como valores, podemos bien estar cuantificando variables claramente observables o más bien variables latentes. En el primer caso estamos haciendo una medida directa, mientras que en el segundo la medida es indirecta. Por ejemplo, medir la distancia recorrida por un objeto entre el punto a y el b se trata de una cuantificación directa de la variable *recorrido*, independientemente de si la unidad es metros, kilómetros o millas. Pero

medir la variable *espiritualidad*, que no puede ser observada directamente, tal vez requiera que contabilicemos un fenómeno observable que consideramos relacionado con ella, como la cantidad de veces que el sujeto visita un templo. Estas medidas indirectas deben ser elegidas con cuidado y con frecuencia generan dudas sobre si realmente apuntan a medir el constructo (¿es posible tener una alta espiritualidad sin asistir nunca a un templo?).

La construcción de un instrumento que permite medir magnitudes es conocida como “escalamiento” (*scaling* en inglés), el cual es “el proceso de acceder introspectivamente a la información de la magnitud y vincularlo con una escala numérica o de otro tipo” (Boring, 2004, p. 22). La cuantificación de magnitudes es central en las subespecialidades psicológicas de la psicofísica y la psicometría (Muñiz, 1998). Una diferencia crucial entre ambas es la claridad y determinación del estímulo: la psicofísica mide reacciones a estímulos físicos (es decir, cuantifica magnitudes perceptuales, como la respuesta a la luz, sonidos, etc.), mientras que la psicometría cuantifica variables psicológicas en que los estímulos son más complejos y difíciles de aislar (es decir, mide magnitudes subjetivas, por ejemplo, ansiedad, locus de control, motivación, etc.), (Boring, 2004). A lo largo de la historia de estas disciplinas también se han generado diferencias en cuanto a la medición en sí, lo cual revisaremos de inmediato.

2. EL ESCALAMIENTO PSICOFÍSICO COMO INICIO DE LA MEDICIÓN EN PSICOLOGÍA

La cuantificación de magnitudes en psicología comenzó a hacerse desde el área de la psicofísica, con el trabajo pionero de Weber y especialmente de Fechner, alrededor de 1860 (Muñiz, 1998; Brenan, 1999; Boring, 2004; Sáiz, 2009). Estas investigaciones asociaban las variaciones en la intensidad de un estímulo con la percepción por parte del sujeto. Por ejemplo, se incrementaba en proporciones iguales y sucesivas la presión de un objeto sobre la piel, y se medía si el sujeto percibía variaciones. Fechner encontró que mientras el estímulo variaba en proporción aritmética (es decir, de uno en uno), la sensación variaba en proporción geométrica (es decir, multiplicándose por un factor), a lo cual se le conoció como Ley de Fechner, o Ley de Weber-Fechner.

Fechner propuso la medición de la sensación mediante los denominados “métodos psicofísicos”, que permitían cuantificar la sensibilidad diferencial. Estos son:

- a. *Método de las mínimas diferencias perceptibles, o de los límites*: Se presentaban dos estimulaciones (por ejemplo, dos puntos de presión sobre la piel, ejercidos mediante un compás de dos puntas), una de las cuales se mantenía fija mientras que la otra se iba alejando. El sujeto debía decir cuándo las percibía como estímulos diferentes y cuándo le parecían un estímulo único.
- b. *Método de los casos verdaderos o falsos, también llamado de estímulos constantes*: Como en el anterior, uno de los estímulos se mantenía constante, mientras el otro se iba difiriendo. El sujeto debía decir si se trataba del mismo estímulo o no.
- c. *Método de los errores promedio o de ajuste*: Se mostraba un estímulo patrón y el sujeto debía ajustar los otros estímulos hasta que, a su juicio, coincidieran con él (Sáiz, 2009).

Alrededor de 1953 Stevens replicó estos experimentos pidiendo a los sujetos que asignaran un valor a su sensación, y consiguió que cuando el estímulo se incrementa en proporción aritmética la sensación lo hace de acuerdo con una función exponencial (Brenan, 1999; Boring, 2004; Sáiz, 2009). Lo fundamental de los trabajos de Stevens es su propuesta de superar el requisito de isomorfismo, aceptando que puede haber diferentes reglas para la medición de las magnitudes (Muñiz, 1998), lo cual abrió el camino a la construcción de escalas más simples y manejables.

3. PROPUESTAS DEL ESCALAMIENTO PSICOLÓGICO

El principal problema que presenta la psicología al nivel de medición es el hecho de que las variables no son directamente observables, al contrario de lo que ocurre en las ciencias naturales (Muñiz, 1998). Debido a esto la cuantificación con mucha frecuencia debe inferirse a partir de ciertos correlatos que resultan indicadores de la variable en sí. Por ejemplo, el nivel de sed de un perro puede inferirse a partir de su salivación o de la cantidad de agua que bebe; la motivación de un maratonista se infiere a partir del número de horas que entrena, etc. Según Cook y Seltiz (1976), algunos de los correlatos utilizables en psicología son:

- a. Medidas en que las inferencias se obtienen de autoinformes sobre creencias, sentimientos, conductas, etc.
- b. Medidas en donde las inferencias se hacen a partir de la observación de la conducta manifiesta.

- c. Medidas en que las inferencias se obtienen de la reacción individual a un estímulo, o de la interpretación de estímulos estructurados parcialmente.
- d. Medidas en que las inferencias se extraen de la ejecución de tareas objetivas.
- e. Medidas en que las inferencias se obtienen de reacciones fisiológicas al objeto actitudinal o a representaciones de él.

El escalamiento psicofísico relaciona estímulos físicos con percepciones del sujeto, pero el individuo es mucho más que un conjunto de respuestas ante el mundo material, lo que hace que las propuestas y métodos de la psicofísica resultan limitados para explicar problemas más complejos. La psicometría inicial surgió como una extensión de los logros del escalamiento psicofísico, es decir, que se intentó la construcción de escalas para variables psicológicas o psicosociales (no físicas), llamando a esto escalamiento psicológico.

Torgerson (1958, c.p. González, s.f.) propone una clasificación para discriminar entre distintos métodos de escalamiento. Un primer nivel de clasificación permite diferenciar entre métodos unidimensionales y métodos multidimensionales:

- a. *Métodos unidimensionales*: Asumen que la variable se comporta como un continuo lineal que realiza un recorrido de menor a mayor.
- b. *Métodos multidimensionales*: La variable puede ser una combinación de diferentes subvariables o distintas dimensiones, que finalmente resultarán agregadas.

En cuanto a la atribución de la variabilidad que dan los sujetos frente a las respuestas, Torgerson (1958, c.p. González, s.f.) clasifica los métodos de escalamiento psicológico de la siguiente forma:

- a. *Métodos centrados en el estímulo o métodos comparativos*: Pretenden la ordenación de estímulos en un supuesto continuo subyacente. Los sujetos responden para lograr esta ordenación de estímulos, y luego los estímulos pueden permitir evaluar al sujeto individual. Se producen luego dos subclasificaciones:
 - *Métodos de juicio cuantitativo*: Se considera exclusivamente la medida central del juicio.

Método de estimación subjetiva: El sujeto asigna un valor numérico subjetivo a un estímulo dado (esta subclasificación es un aporte de Garner & Creelman, 1976).

Método de fraccionamiento: Se le pide a un sujeto que elabore una magnitud de estímulo que sea una fracción específica de otro, o que sea múltiplo de otro estímulo dado. Por ejemplo, producir un sonido el doble de fuerte (Garner & Creelman, 1976).

Método de equisección: Existe un estímulo físico verdadero que funciona como un continuo y este es dividido en intervalos iguales (por ejemplo, intensidad de un sonido o concentración de azúcar de una bebida), (Garner & Creelman, 1976).

Método de suma constante: Se comparan dos estímulos, pero en lugar de identificar uno mayor y uno menor se proporciona la respuesta en una escala de razón que debe sumar 100. Por ejemplo, se dirá que el estímulo A obtiene 70, y el B 30, queriendo esto decir que el estímulo A es mayor que el B y en determinada medida. El método se trabaja a través de diferencias logarítmicas (Garner & Creelman, 1976).

- *Métodos basados en la variabilidad de los juicios:* Se incorpora la medida de variabilidad en los juicios, es decir, la varianza de las respuestas de los sujetos (métodos de sensibilidad diferencial, métodos thurstonianos: comparaciones pareadas e intervalos aparentemente iguales)

- b. *Métodos centrados en el sujeto:* Asumen que la variabilidad es producto de diferencias individuales de la capacidad que tienen los sujetos al enfrentarse a un estímulo y busca ordenar diferentes sujetos dentro de un continuo, asumiendo que los estímulos son equivalentes.
- c. *Métodos de respuesta:* Ordenan simultáneamente a los sujetos y a los estímulos.

Estos distintos enfoques han producido escalas diferentes, muchas de las cuales son aun ampliamente utilizadas. A continuación pasaremos a explorar cómo se elaboran escalas según los métodos más reconocidos.

3.1. MÉTODOS UNIDIMENSIONALES CENTRADOS EN EL ESTÍMULO: MÉTODOS THURSTONIANOS

Se trata de dos procedimientos (aunque al hablar de Thurstone suele considerarse solo el segundo): el *Método de Comparaciones Pareadas* y el *Método de Intervalos Aparentemente Iguales*.

3.1.1. *Método de Comparaciones Pareadas*

Louis Leon Thurstone fue uno de los pioneros en el desarrollo del escalamiento psicológico, hacia el año 1927 (Brenan, 1999; Sáiz, 2009). El método de Thurstone plantea un tratamiento tan riguroso de los estímulos psicológicos como para lograr que estos se comporten como lo hacen los estímulos psicofísicos, basándonos en la percepción de los sujetos y no en la medición del estímulo (Thurstone, 1928; Elejabarrieta e Iñiguez, 1984).

Según un excelente ejemplo de Bucla & Sierra (1997), para ordenar una serie de objetos según un continuo vinculado a la variable peso (es decir, para organizarlos desde el más liviano hasta el más pesado) tenemos dos alternativas: medir los objetos con una balanza (y si eso fuera posible, sería el método más exacto) y luego ordenarlos, o pedir a un grupo de sujetos que actúen como jueces y ordenen los objetos según su juicio particular sobre el peso. Este orden solo puede lograrse comparando pares de objetos (por ejemplo, b es más pesado que a). Para lograr un continuo de varios niveles ubicados a la misma distancia es necesario que múltiples sujetos hagan múltiples comparaciones de cada par de estímulos (A con B, A con C, C con D, B con C, y así hasta que todos evalúen todos los pares). Debido a esto, al método de Thurstone también se le conoce como *Método de Comparaciones Pareadas* o *Escala Diferencial de Thurstone*.

Numerosas comparaciones pareadas hechas por múltiples sujetos permitirán finalmente crear una escala que represente el continuo unidimensional de la variable estudiada. Solo resta lograr que los “peldaños” de esa “escalera” (es decir, la distancia de un estímulo a otro) sea de la misma distancia para cada caso, lo cual se logra comparando las medias de los estímulos originales y eligiendo aquellas que estén a distancias similares. Esta escala es de por sí interesante en la comprensión del fenómeno y al ser aplicada sirve también para evaluar la actitud de un nuevo sujeto que no haya participado en el análisis original que dio lugar a la escala.

La deficiencia de este procedimiento estriba fundamentalmente en su dificultad: es necesario que cada sujeto compare de a pares decenas de estímulos (A con B, A con C, A con D, B con C, etc.), lo cual es un procedimiento largo y engorroso (si tenemos apenas diez estímulos, cada sujeto debe hacer cuarenta y cinco comparaciones). Por otro lado, debido a la fatiga y la saturación, no es muy alto el número de comparaciones que un individuo puede hacer. Esto juega en contra de la técnica, pues la probabilidad de obtener ítems ubicados a distancias iguales aumentaría a medida que aumentamos la cantidad de ítems probados.

3.1.2. Método de Intervalos Aparentemente Iguales

Debido a que el *Método de Comparación Pareada* es largo y engorroso (por el número de pares que deben ser comparados), Thurstone desarrolló una alternativa más sencilla, conocida como *Método de Intervalos Aparentemente Iguales*. En él se recurre también a la opinión de jueces, los cuales tienen que ubicar un determinado número de frases dentro de las diferentes categorías de una escala, que puede ir del 0 al 10 (once montones diferentes), lo cual produce un valor para cada frase y cada sujeto. El tratamiento estadístico en adelante es simple: se calcula la media y se seleccionan once estímulos (uno correspondiente a cada etiqueta del continuo), ubicados a igual distancia unos de otros.

Una escala de Thurstone es, finalmente, un conjunto de frases ordenadas, y se supone que el procedimiento garantiza que la distancia entre una frase y la siguiente es la misma. La escala se convierte en un compendio altamente estructurado de estímulos que elicitán respuestas simples y fáciles de medir: la respuesta de los sujetos es un simple “sí” o “no” expresando, su acuerdo o desacuerdo con la frase. Veamos algunos ítems de una escala Thurstone (Elejabarrieta & Iñíguez, 1984).

Pienso que la Iglesia es un parásito de la sociedad

La Iglesia representa espíritu superficial, hipocresía y prejuicios

Siento la necesidad de una religión, pero no encuentro lo que quiero en ninguna Iglesia

Creo que la Iglesia es una acción potente para promover la virtud en la fe individual y social

Como vemos, las actitudes hacia la Iglesia se ordenan desde la más negativa hasta la más positiva. La escala ha sido desarrollada de forma que el intervalo percibido entre una oración y la siguiente sea siempre el mismo.

Como podrá notarse, los métodos de Thurstone guardan estrecha similitud con las bases del escalamiento psicofísico:

- a. Se trata de estímulos ordenados de forma creciente (análogo a la función monótona creciente en la presentación de estímulos del escalamiento psicofísico, por ejemplo, aumentar sucesivamente la presión sobre la piel).
- b. El método permitiría que el intervalo entre estímulos fuera constante (del mismo modo en que, en el escalamiento psicofísico, se mantiene constante el delta entre un nivel dado del estímulo y el nivel siguiente).
- c. Se presupone coherencia en el individuo, de modo que al responder positivamente a un estímulo debe haber respondido de la misma forma a todos los estímulos precedentes (lo que daría origen, posteriormente, a la llamada forma característica de la curva, en la Teoría de Respuesta al Ítem).

3.1.3. Escalas Thurstone y Teoría de Respuesta al Ítem (TRI)

Resaltemos tres aspectos en el modo general de funcionamiento de una escala Thurstone:

- a. Se trata de una muestra única de conducta: la escala es considerada como un todo y no hay medidas repetidas de la muestra de conducta. Por ejemplo, si hablamos de actitudes hacia los extranjeros, utilizaremos la escala Thurstone aplicada una sola vez, considerando a todos sus ítems de manera conjunta y como parte de una única escala. La puntuación en la escala es la puntuación total del sujeto.
- b. Se supone que el estímulo (es decir, las frases ordenadas) se comportan como un continuo unidimensional, exactamente de la misma forma en que se comporta la intensidad de la luz mostrada en un experimento psicofísico. Por ejemplo, en una prueba de potencia, quienes responden a los ítems más difíciles seguramente responderán también a los más fáciles.

- c. Las frases están ordenadas de acuerdo con un gradiente, de modo que un sujeto relativamente coherente debe parar de responder afirmativamente en un determinado lugar.

Estas son características que asemejan las escalas Thurstone con las investigaciones psicofísicas en las cuales se inspiraron y que también las aproximan a los modelos psicométricos de la Teoría de Respuesta al Ítem, en contraposición a la Teoría Clásica de Test (Thurstone, 1928; Atorresi, Lozia, Abal, Galibert & Aguerri, 2009; Sánchez, s.f.). Según Muñiz (1998), la Teoría de Respuesta al Ítem (TRI) fue desarrollada primeramente en los trabajos de Rash en 1960 y de Lord & Novick en 1968. Una revisión más profunda de la TRI excede los límites de este trabajo.

3.2. MÉTODOS UNIDIMENSIONALES CENTRADOS EN LOS SUJETOS

Los métodos de Thurstone producen escalas donde los estímulos están ordenados de acuerdo con la organización preferente de los sujetos (es decir, una organización promedio) y debido a que se centran en la ordenación de estímulos se conocen como métodos centrados en el estímulo. Veamos ahora la escala Likert, que es el principal exponente de desarrollos que prestan menos atención al modo en que se ordenan los estímulos y se centra fundamentalmente en la medición de los sujetos.

3.2.1. *Escala Likert*

En la literatura es común la referencia a que la escala Likert surge como respuesta a las dificultades para la construcción de escalas Thurstone. Likert desarrolló su técnica alrededor de 1932. Esta escala pretende que cada ítem constituya una medida repetida de muestras de conducta, lo que quiere decir que los sujetos son evaluados a través de estímulos considerados equivalentes, los cuales son presentados en numerosas ocasiones (Muñiz, 1998; Sánchez, s.f.).

Lo que hace la escala Likert es trasladar la medición de las diferencias de intensidad hacia las opciones de respuesta, evitando el paso de calibrar la intensidad de las frases. De esta manera, una escala Likert se compone de un conjunto de frases no calibradas previamente, ante las cuales el sujeto da una respuesta con un gradiente de intensidad. Un ítem típico de una escala Likert sería, por ejemplo:

Me gusta ir a la iglesia cada vez que puedo

1. Totalmente de acuerdo
2. De acuerdo
3. Ni de acuerdo ni en desacuerdo
4. En desacuerdo
5. Totalmente en desacuerdo

Otro ítem de la misma escala podría ser:

Considero que ir a la iglesia es la obligación de todo creyente

1. Totalmente de acuerdo
2. De acuerdo
3. Ni de acuerdo ni en desacuerdo
4. En desacuerdo
5. Totalmente en desacuerdo

Como base para la construcción de una escala Likert, existen ciertos supuestos:

- Muestreo sucesivo y equivalente:* Los ítems son considerados medidas equivalentes o muestras independientes y repetidas de conducta.
- Unidimensionalidad:* se supone que todos los ítems miden una y solo una variable, evaluada en una y solo una dimensión. Esta unidimensionalidad tiene exigencias, muchas de las cuales son supuestos básicos de la Teoría Clásica de Tests (Muñiz, 1998; Sánchez, s.f.):

Los sujetos deben exhibir coherencia entre sus respuestas (lo que se expresa estadísticamente como una alta correlación intratest). A la coherencia interna del test se refieren varias de las medidas de confiabilidad (como el conocido alpha de Cronbach).

Debe existir una alta correlación entre cada ítem y la medida global del test.

El análisis factorial debe mostrar la agrupación de todos los ítems cerca de un solo eje principal.

- c. *Nivel de medida*: Aunque el nivel de medida de los ítems Likert puede ser considerado como máximo, de intervalo, la suma de puntajes genera tanta dispersión que permite que la puntuación total del test sea considerada, usualmente, de razón, lo cual permite la aplicación de estadísticos más sofisticados (Carifio & Perla, 2007).

Es común la utilización de “preguntas tipo Likert” en la medición de opinión pública y la investigación de mercado, por ejemplo, en la investigación por encuestas. Sin embargo, es incorrecto llamar a estas preguntas “de escala Likert”, pues hay infinidad de supuestos relacionados con la construcción de auténticas escalas Likert que no se cumplen en la investigación por encuestas. En particular, es importante señalar que la escala solo tiene sentido cuando se agregan (suman) las puntuaciones en los diferentes ítems que la componen. Los ítems, vistos por separado, no tienen valor, al menos dentro del quehacer de la psicometría. En tal sentido, la escala y la prueba o test son lo mismo.

Mucha de la literatura alrededor de la construcción de escalas Likert tiene que ver con la elaboración de las opciones de respuesta. Preguntas usuales son: ¿Cuál es la cantidad adecuada de opciones de respuesta? ¿Deben utilizarse etiquetas en cada opción o pueden usarse simplemente números? ¿Si se usan números, deben usarse solo positivos o positivos y negativos? (Chang, 1994; Cummins & Gullone, 2000; Garland, 1991; Jacoby & Mattel, 1971). A continuación veremos diferentes opciones de respuesta para un mismo ítem (inspirado en Dagostín, 2005, con aportes personales):

Cuadro 1.

Alternativas de presentación de un ítem de una escala

Escala impar, cinco puntos, numeración directa	Es importante tener creencias religiosas 1. Totalmente de acuerdo 2. De acuerdo 3. Ni de acuerdo ni en desacuerdo 4. En desacuerdo 5. Totalmente en desacuerdo
--	---


→ Continuación

Cuadro 1. *Alternativas de presentación de un ítem de una escala*

Escala impar, siete puntos, numeración directa	<p>Es importante tener creencias religiosas</p> <ol style="list-style-type: none"> 1. Totalmente de acuerdo 2. De acuerdo 3. Algo de acuerdo 4. Ni de acuerdo ni en desacuerdo 5. Algo en desacuerdo 6. En desacuerdo 7. Totalmente en desacuerdo
Escala impar, cinco puntos, numeración inversa	<p>Es importante tener creencias religiosas</p> <ol style="list-style-type: none"> 5. Totalmente de acuerdo 4. De acuerdo 3. Ni de acuerdo ni en desacuerdo 2. En desacuerdo 1. Totalmente en desacuerdo
Escala impar, cinco puntos, numeración directa, el punto medio está representado como un 0	<p>Es importante tener creencias religiosas</p> <ol style="list-style-type: none"> 2. Totalmente de acuerdo 1. De acuerdo 0. Ni de acuerdo ni en desacuerdo -1. En desacuerdo -2. Totalmente en desacuerdo
Escala impar, cinco puntos, numeración directa, solo se explicitan las etiquetas de los extremos	<p>Es importante tener creencias religiosas</p> <ol style="list-style-type: none"> 1. Totalmente de acuerdo 2. 3. 4. 5. Totalmente en desacuerdo
Escala impar, tres puntos, sin numeración, pero con orden directo (necesaria ponderación posterior)	<p>Es importante tener creencias religiosas</p> <ol style="list-style-type: none"> a) Totalmente de acuerdo b) Ni de acuerdo ni en desacuerdo c) Totalmente en desacuerdo

→ Continuación

Cuadro 1. *Alternativas de presentación de un ítem de una escala Likert*

Escala par, cuatro puntos, sin numeración, con orden directo (necesaria ponderación posterior). Expertos sugieren que la ausencia de punto medio evita respuestas “cómodas” y maximiza la varianza	Es importante tener creencias religiosas a) Totalmente de acuerdo b) De acuerdo c) En desacuerdo d) Totalmente en desacuerdo
Escala gráfica impar, tres puntos, sin numeración, con orden directo (necesaria ponderación posterior). Muy útil para el trabajo con niños	Es importante tener creencias religiosas 
Escala gráfica sin cortes, sin numeración (se obtiene el número midiendo la distancia desde el extremo izquierdo hasta la marca del sujeto)	Es importante tener creencias religiosas Totalmente de acuerdo ————— Totalmente en desacuerdo
Escala gráfica de siete puntos, con cortes, sin numeración (el número está implícito en cada corte, pero el sujeto no lo ve)	Es importante tener creencias religiosas Totalmente de acuerdo Totalmente en desacuerdo
Escala gráfica de siete puntos, con cortes, con numeración	Es importante tener creencias religiosas Totalmente de acuerdo Totalmente en desacuerdo 1 2 3 4 5 6 7

Estas son solo algunas de las opciones más usadas. Carifio & Perla (2007) mencionan estudios que evalúan la coherencia en las medidas de actitudes logradas según diferentes modelos de ítem, y consiguen que son muy similares, sin importar el modelo del ítem. Incluso hay bastante correlación entre preguntas abiertas codificadas y medidas logradas mediante escalas. Cummins & Gullone (2000) votan por el uso de escalas de tres puntos, que no muestran diferencias sustanciales en términos de la varianza en relación con las de cinco puntos. Hernández, Espejo, González & Gómez (2001) apuntan a utilizar escalas con un punto medio.

Un aspecto importante es el nivel de medición de cada modelo de ítem. Obviamente, algunos de ellos (como el modelo gráfico de los rostros sonrientes) de ninguna manera alcanzan un nivel de intervalo-razón, ubicándose si acaso en un nivel de medición ordinal, por lo cual ni siquiera deberían ser sumados para componer un índice global de la prueba. Según algunos autores (Knapp, 1990; Carifio & Perla, 2007; Norman, 2010), la evidencia muestra que las pruebas estadísticas paramétricas son poco sensibles al uso de niveles de medición de razón *versus* intervalo, de modo que es razonable suponer que no existirán sesgos importantes trabajando con los puntajes agregados a partir de ítems con un razonable número de opciones de respuesta (la literatura habla de 7 o más opciones de respuesta por ítem para esperar que el puntaje total del test tenga cualidades de intervalo-razón). En ocasiones (como en el ejemplo gráfico de las “caritas”) es un problema de pureza estadística *versus* ecología del instrumento. Vale señalar que para casos como este, otros autores (Göb, McCollin & Ramalhoto, 2007) abogan por utilizar análisis propiamente desarrollados para datos ordinales.

3.3. MÉTODOS UNIDIMENSIONALES DE ORDENAMIENTO SIMULTÁNEO DE SUJETOS Y RESPUESTAS

Hemos visto que la meta de las escalas thurstonianas se estructuran alrededor de presentar resultados sobre el orden promedio de los estímulos y que la escala Likert intenta fundamentalmente la medición de sujetos. Veremos ahora algunos métodos que abordan simultáneamente la medición de ítems y de individuos.

3.3.1. Escalograma de Guttman

Como puntualizan Dotson & Sommers (1976), lo característico de la escala de Guttman es que entre los reactivos que forman la escala existe una relación que permite que una persona que concuerda con un reactivo que ocupa una posición escalar dada, también estará de acuerdo con todos los reactivos que estén por debajo de ese en la escala. Este requisito de coherencia (que no existía ni siquiera en los métodos thurstonianos) garantiza que si diez individuos respondieron afirmativamente a tres ítems, hayan sido siempre los mismos tres ítems (Dotson & Summers, 1976).

Hay varias técnicas para elaborar un escalograma de Guttman, pero en todas ellas el objetivo fundamental es procurar que los ítems estén estrictamente organizados de acuerdo con un gradiente.

- a. Una técnica muy simple consiste en imprimir cada frase en una tablilla o tarjeta, y luego entregar estas a los sujetos para que las ordenen. Se puede intentar lograr un acuerdo consensuado o promediar las posiciones y seleccionar las que presenten menor variabilidad (Guttman, 1976).
- b. Otra técnica parte de utilizar cada frase como el pie de una escala tipo Likert, con cinco alternativas de respuesta. Luego se toma la distribución de frecuencias acumuladas de cada opción de respuesta, desde la más positiva hasta la más negativa, y se selecciona un punto de corte arbitrario para elegir a los mejores ítems y su orden. Se seleccionan frases con suficiente brecha de respuesta entre sí, lo cual disminuye la probabilidad de una inversión en el orden para algún sujeto particular (Guttman, 1976).
- c. Una tercera técnica parte de reactivos realmente dicotómicos, por ejemplo, con opciones de respuesta “sí” y “no”, que se escalan de acuerdo con la frecuencia de sujetos que responden afirmativamente. Nuevamente, se seleccionará un conjunto de frases con suficiente brecha entre sí, para evitar inversiones en sujetos particulares (Guttman, 1976).

Un ejemplo de un ítem Guttman puede ser tan sencillo como el siguiente:

- | |
|---|
| <ol style="list-style-type: none">4. Me siento motivado a trabajar 8 horas seguidas3. Me siento motivado a trabajar 6 horas seguidas2. Me siento motivado a trabajar 4 horas seguidas1. Me siento motivado a trabajar 2 horas seguidas |
|---|

Para verificar la consistencia del listado de reactivos se realizan tablas de doble entrada 2x2 con cada par de ítems, y se espera que la proporción de casos de inversión sea inferior al 10% (casos contenidos en la diagonal secundaria, es decir, la suma de las celdillas +- y -+), como puede verse a continuación.

Cuadro 2.

Ejemplo de presentación de acuerdos e inversiones en un buen ítem para una escala Guttman

	Sí (ítem 3) +	No (ítem 3) -
Sí (ítem 4) +	45	8
No (ítem 4) -	3	23

Suponiendo que se lograra un escalograma perfecto, los patrones de respuesta (es decir, la forma que adquiere la respuesta de cada individuo) pueden ser perfectos o presentar error. Un patrón perfecto de respuesta es aquel que ocurre de acuerdo con el supuesto escalar que indica que la secuencia de respuestas positivas debe cortarse en algún punto y dar paso a una secuencia de respuestas negativas, sin que se produzcan inversiones. En un escalograma de cuatro frases ocurren cinco patrones puros: +++, ++-, +-+, +-- y ----.

Por supuesto, es muy improbable que todos los sujetos a los que se le aplica el escalograma final logren una coherencia tal como la esperada. Los sujetos que presentan patrones no puros (o errores) son reasignados a un patrón puro de acuerdo con un criterio denominado “de error mínimo”, lo que significa que se les reasigna al patrón que presente menos discordancias o errores con respecto a la respuesta particular que dieron. Por ejemplo, un sujeto que haya respondido ++++ será reasignado al patrón ++--. No siempre la reasignación resulta clara.

3.3.2. Escala de distancia social de Bogardus

La escala de distancia social de Bogardus fue desarrollada en 1925 con la intención de medir prejuicios raciales, de modo que no evalúa directamente la actitud, sino disposición a cierto tipo de proximidad social. Los ítems de una escala de distancia social están escalados al igual que los de la escala de Guttman, de modo que responder afirmativamente a un ítem implica (al menos teóricamente) responder afirmativamente a ítems de intensidad inferior. El ejemplo clásico de Bogardus, referido a grupos raciales, puede ser el mejor:

1. ¿Aceptaría que personas de esa raza entraran solo como turistas a este país?
2. ¿Aceptaría que vivieran en este país?
3. ¿Los aceptaría como compañeros de trabajo?
4. ¿Los aceptaría como vecinos en su calle?
5. ¿Sería amigo de personas de esa raza?
6. ¿Aceptaría que entraran a su familia casándose con un familiar suyo?

Como vemos, los ítems se escalan de forma tal que el primero es que implica menor proximidad social y el último representa el mayor grado de proximidad. Teóricamente se produce un “punto de quiebre” de forma que responder negativamente al ítem 4, por ejemplo, implicará una respuesta negativa en ítems de mayor peso (es decir, el 5 y el 6).

3.3.3. Técnica de despliegue de Coombs

La técnica de Coombs permite ubicar a los sujetos dentro de un continuo, pero también permite medir el supuesto de que los diferentes tramos de este continuo son de igual distancia o no.

Se denomina técnica de despliegue porque en algún momento el sujeto “dobla” la escala original a partir de su preferencia de estímulos. Por ejemplo, la escala original presenta los estímulos a, b, c, d, e y f. Si el sujeto prefiere el estímulo c, es natural que a partir de él ordene su preferencia en d, b, e, a y f, generando una especie de dos subseries (una hacia la izquierda, con los estímulos que “prefiere menos” y hacia la derecha, con los que prefiere realmente).

El escalamiento se produce comparando la probabilidad de una serie de patrones teóricamente posibles (asumiendo igualdad entre los intervalos) con la serie de patrones de ordenación que efectivamente ocurre, lo cual indica que los intervalos no son iguales. Al igual que en la técnica de Guttman, el punto crítico de este método es la consistencia entre las respuestas de los sujetos.

3.4. MÉTODOS MULTIDIMENSIONALES DE RESPUESTA

Hemos visto que los métodos unidimensionales parten del supuesto de que la escala mide exclusivamente una dimensión del constructo evaluado, lo cual es fundamental, sobre todo, para aquellos que ordenan estímulos. Los

métodos multidimensionales, por el contrario, admiten que el constructo posee más de una dimensión. Dichas dimensiones están definidas tanto desde el punto de vista de la teoría como desde la perspectiva del manejo de los resultados.

El ejemplo más claro de un método multidimensional de medida es el diferencial semántico.

3.4.1. Diferencial semántico de Osgood

El método de diferencial semántico de Osgood se considera una técnica específica para la medición de actitudes. El modelo clásico de DS trabaja con tres factores en los cuales se califica a los objetos actitudinales: evaluación, potencia y actividad (Osgood, Suci & Tannenbaum, 1976). Heise (1976) menciona varias investigaciones fundacionales en las cuales se comprueba que la modelación mediante estas tres dimensiones tiene sentido, incluso transcultural. Para lograr una aproximación intuitiva a estas tres dimensiones, señalaremos algunas definiciones polares (mencionadas también por Heise, 1976):

Evaluación:

bonito-feo, bueno-malo, dulce-ácido, sano-enfermizo.

Potencia:

grande-pequeño, poderoso-impotente, fuerte-débil, profundo-superficial.

Actividad:

rápido-lento, vivo-muerto, ruidoso-quieto, joven-viejo.

Podemos resumir los pasos para elaborar una escala de diferencial semántico (Santoro, 1975):

- a. *Selección de los conceptos*: Debe tratarse con conceptos familiares y relevantes para los sujetos y que estén dentro de su campo de experiencias. De igual forma, debe presentarse un número adecuado de conceptos.
- b. *Selección de las escalas*: La elaboración de las escalas pasa, fundamentalmente, por la selección de los adjetivos polares que definen sus límites. Para ello se sugiere.

- b.1) Presentar a un grupo de sujetos la lista de conceptos a evaluar y pedirles que generen adjetivos respecto a ellos.
- b.2) Pedir que suministren opuestos (antónimos) para cada adjetivo.
- b.3) Determinar el tipo de escala a utilizar (gráfica o numérica, cantidad de intervalos, etc.).
- b.4) Probar las escalas generadas, administrándolas a un grupo de sujetos y correr un análisis factorial para observar la agrupación espontánea de las escalas.
- c. *Seleccionar las mejores escalas tomando en cuenta la carga factorial:*
Heise (1976) sugiere seleccionar tres o cuatro escalas para cada factor, lo cual generaría un instrumento de nueve a doce ítems para cada objeto actitudinal.

A modo de ejemplo, se muestran a continuación algunas de las escalas utilizadas por Santoro (1975) en su trabajo clásico.

Vivo	_ _ _ _ _ _ _ _ _	Muerto
Negro	_ _ _ _ _ _ _ _ _	Blanco
Cobarde	_ _ _ _ _ _ _ _ _	Valiente
Estable	_ _ _ _ _ _ _ _ _	Inestable

La aproximación multifactorial mediante tres dimensiones (evaluación, potencia, actividad) genera una medida de la actitud que es a su vez tridimensional. En diferencial semántico se genera un puntaje por ítem o por factor (E-P-A), pero no se reporta un puntaje general para el objeto evaluado.

Aunque cada objeto actitudinal puede estar conceptualmente ligado a adjetivos particulares, se ha hecho un esfuerzo internacional por definir adjetivos polares que permitan una buena evaluación de gran variedad de objetos actitudinales, constituyendo una especie de “banco de escalas”. Esta capacidad de utilizar escalas similares para diferentes objetos actitudinales es señalada como una de las ventajas de la técnica de diferencial semántico.

4. UN APÉNDICE: OPERACIONES MATEMÁTICAS EN EL ANÁLISIS DE ESCALAS

Usualmente se considera que hay cuatro niveles de medición: nominal, ordinal, intervalo y razón; los dos primeros llamados “cualitativos” y los dos últimos “cuantitativos”. El nivel de medición de intervalo requiere tramos iguales en el continuo de medición, para garantizar lo cual se requeriría un análisis matemático de la distancia entre los ítems de las escalas (Cañadas & Sánchez, 1998). Adicionalmente, para hablar de un nivel de medición de razón es necesario que exista un cero absoluto, tanto en la variable subyacente como en su escala de medida. Esto generalmente no ocurre en psicología (por ejemplo, nunca hablamos de “inteligencia nula” y los tests de inteligencia nunca arrojan puntajes de 0 para determinado individuo).

Debe recordarse que con los métodos de escalamiento el nivel de medición de razón se logra con la prueba vista como un todo y no en la consideración individual de los ítems. Considerado de esta manera, el instrumento no solo tiene un nivel de dispersión de los más elaborados, sino que pasa a ser, aproximadamente, continuo debido a la gran dispersión en los posibles puntajes totales, a pesar de que la medida original de los ítems era indudablemente discreta. (En relación con la discusión sobre la continuidad de las medidas psicofísicas, se produjo un largo debate, que involucró pruebas matemáticas y cuya historia puede verse con más detalle en Boring [2004].) Otra ventaja de asumir que se ha logrado un nivel de medición de intervalo-razón, es que eso nos permite realizar operaciones matemáticas complejas y así acceder a análisis más potentes y de mayor profundidad. Lograr un nivel de intervalo-razón y suponer continuidad tanto en la variable subyacente como en la escala hace posible transformar de múltiples formas los datos originales. Veamos un par de ejemplos:

- a. El producto vectorial, que implica la multiplicación de dos escalas.
- b. El producto escalar, que significa la multiplicación de la medida de la escala por alguna constante.

4.I. UN EJEMPLO DE PONDERACIÓN PRODUCTO VECTORIAL: MODELO EXPECTATIVA-VALOR DE LA TEORÍA DE LA ACCIÓN RAZONADA DE FISHBEIN Y AJZEN

La Teoría de la Acción Razonada, planteada en 1967 por Fishbein y Ajzen, parte del supuesto de que los seres humanos son esencialmente racionales, y busca comprender y predecir probabilísticamente la emisión de conductas a

partir de la evaluación de las mismas por parte de los sujetos (Reyes, 2007; Brenan, 1999; Sáiz, 2009). Por ejemplo, según los planteamientos de estos autores, la actitud de una persona hacia un objeto lograría ser medida con exactitud mediante la sumatoria de la multiplicación de la evaluación de cada atributo asociado con un objeto X, por la probabilidad subjetiva de que el objeto tenga tal atributo. Este procedimiento implica la medición de cada uno de los sujetos en dos escalas: una escala de creencia y una escala de evaluación de la creencia. Por ejemplo:

a) Hacer bien a los demás traerá resultados positivos

(*evaluación conducta-resultado*)

1. Completamente de acuerdo
2. De acuerdo
3. Ni de acuerdo ni en desacuerdo
4. En desacuerdo
5. Completamente en desacuerdo

b) Esto ocurre...

(*evaluación de la fuerza de la creencia*)

1. Pocas veces
2. Algunas veces
3. Casi siempre
4. Siempre

Dado que los dos ítems están enlazados, para obtener la evaluación de un sujeto en relación con esta actitud es necesario multiplicar el puntaje que dio en la *evaluación conducta-resultado* por el puntaje que dio en la *evaluación de la fuerza de la creencia*. Los puntajes simples de los ítems se convierten en factores de ponderación, también llamados de expansión.

4.2. UN EJEMPLO DE PONDERACIÓN PRODUCTO ESCALAR: PONDERACIÓN DE ESCALAS DE INGRESOS

Este ejemplo se deriva de una práctica cotidiana en investigación de mercado, que consiste en determinar el ingreso promedio de un determinado grupo a partir de las respuestas dadas a un ítem cuyo nivel de medición es ordinal.

Un ítem típico para levantar información sobre ingresos podría ser el que se muestra a continuación. Se trata de una pregunta con nivel de medición ordinal debido a que los rangos son desiguales y que no es posible determinar

la existencia de un 0 absoluto (que quedaría confundido dentro del nivel más bajo de la escala, la alternativa).

Por favor, dígame en cuál de los siguientes rangos se ubica el ingreso de su grupo familiar:

- a) Menos de Bs. 3.500
- b) Entre Bs. 3.501 y Bs. 5.000
- c) Entre Bs. 5.001 y Bs. 10.000
- d) Más de Bs. 10.001

Ítems de este tipo suelen reportarse a través de frecuencias y porcentajes, lo cual es un tratamiento adecuado para niveles de medida ordinales. La ponderación da, sin embargo, información más potente. La práctica común es definir un valor que represente a cada rango. Cuando sea posible, este valor será la marca de clase (punto central del intervalo), que se logra sumando ambos extremos del intervalo y dividiéndolos entre dos (por ejemplo, la marca de clase para el intervalo b es Bs. 4.250,5). Para los intervalos semicerrados (como el a y el d, que tienen límite en uno solo de sus extremos), la marca de clase suele quedar a juicio del investigador (en algunos casos, arbitrario, como en el nivel d de este ejemplo).

Vemos a continuación la media ponderada calculada para este ítem (las frecuencias son ficticias).

Cuadro 3.

Ejemplo de ponderación a partir de una escala ordinal

	Marca de clase	Frecuencias	Marca de clase x Frecuencias
Menos de Bs. 3.500	1.750,0	30	52.500
Entre Bs. 3.501 y Bs. 5.000	4.250,5	328	1.394.164
Entre Bs. 5.001 y Bs. 10.000	7.500,5	432	3.240.216
Más de Bs. 10.001	15.000,0	51	765.000
	Sumatoria	841	5.451.880

Fuente: Elaboración propia.

Para este caso, la media ponderada resultante de la división entre la sumatoria de “Marca de clase x frecuencia” y “frecuencias” es de Bs. 6.482,62. Debido

a las licencias que se han tomado, el promedio de ingresos obtenido por esta vía puede ser sustancialmente diferente al que se lograría preguntando a cada sujeto el ingreso de su familia y calculando posteriormente la media.

4.3. UN EJEMPLO DE ANÁLISIS A PARTIR DE MEDIAS: MATRIZ IPA (IMPORTANCE-PERFORMANCE-ANALYSIS)

Las matrices IPA son muy frecuentes en la investigación de mercado, especialmente en los análisis de satisfacción del cliente. Conuerdan con el modelo de expectativa-valor en relación con la necesidad de medir el atributo en dos dimensiones: la importancia para el cliente *versus* el desempeño de una determinada marca al respecto. En la matriz IPA, cada una de estas evaluaciones (es decir, importancia o desempeño) constituyen un eje y al cruzar ambos ejes se genera un plano cartesiano. El análisis trabaja con los desvíos de los puntajes para producir un dispersograma centrado, lo cual produce cuatro cuadrantes: I (importante-bien evaluado), II (importante-mal evaluado), III (poco importante-mal evaluado) y IV (poco importante-bien evaluado). La decisión gerencial llevaría a enfocar la atención en el cuadrante I.

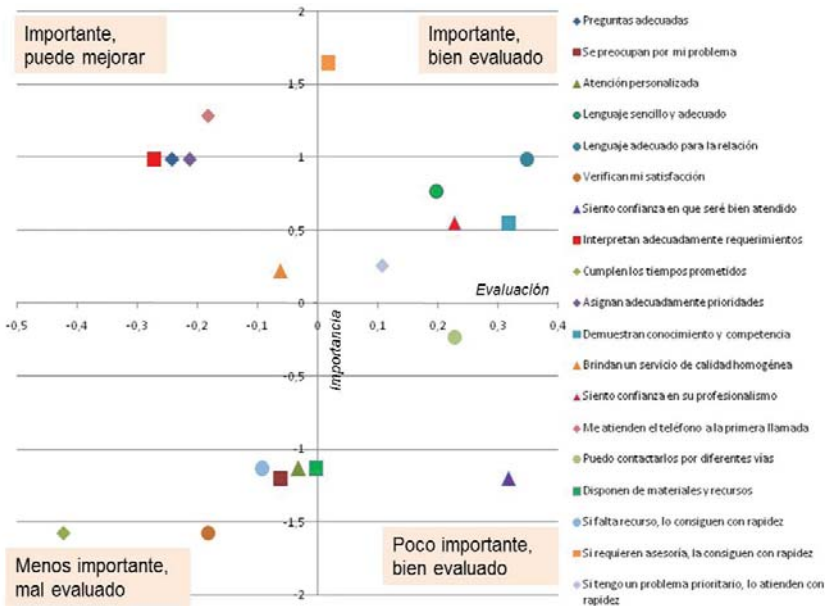


Figura 2. Matriz IPA para un servicio interno de una empresa

EL FUTURO DEL ESCALAMIENTO EN PSICOLOGÍA

La intensa discusión sobre los métodos de escalamiento que se dio alrededor de la década de los cincuenta del siglo xx se ha venido dejándose de lado, y el trabajo del día a día suele echar mano de las escalas más conocidas, sencillas y fáciles de comunicar (en especial las Likert). La escasa literatura reciente sobre el tema de la construcción de escalas continúa trabajando con los métodos tradicionales que hemos revisado, planteando problemas estadísticos y técnicos, tales como el número o calidad de las alternativas y la conveniencia o no de determinado procesamiento de datos. Esto puede deberse a varias causas:

- Boring (2004) indica que se ha producido un cambio histórico de abordaje desde una concepción conservadora (en la cual se buscaba el isomorfismo estímulo-medición, como ya vimos) a una concepción liberal (un punto de vista más flexible y centrado en la aplicación, que incluye medir ciertas dimensiones gracias a su concordancia con otras dimensiones). Respecto a este nuevo enfoque, Muñiz indica que se caracteriza por “reconocer explícitamente el papel que juega la teoría

en la medición, pasando esta a formar parte integral de la teoría” (Muñiz, 1998, p. 16). Esto es llamado “enfoque representacional sobre teoría de la medición”, y se produjo alrededor de 1960.

- b. Boring (2004) señala que la psicología cognitiva (más en boga hoy en día) se enfoca más en los procesos mentales de clasificación que en la estimación de magnitudes, tal vez influida por su raigambre computacional.

Nuestra revisión de la teoría más reciente sobre construcción de escalas parece mostrar algunas líneas de trabajo que ya están presentes y que seguirán desarrollándose en un futuro.

- a. La Teoría de Respuesta al Ítem (TRI), que promete crear instrumentos más eficientes, con mayor potencial discriminativo a través de la aplicación de un menor número de ítems, apoyándose en programas de computadora con algoritmos específicos para tomar decisiones sobre la marcha (Muñiz, 1998; Sánchez, s.f.; Atorresi et al., 2009).
- b. El refinamiento de las escalas producidas dentro de la teoría clásica de construcción de instrumentos psicométricos, ahondando en puntos técnicos como la cantidad de alternativas, la distancia entre los ítems, el uso de análisis estadísticos adecuados, etc. (Knapp, 1990; Cummins & Gullone, 2000; Hernández et al., 2001; Carifio & Perla, 2007; Göb et al., 2007; Norman, 2010).
- c. Un movimiento para aproximar la medición de magnitudes a los desarrollos teóricos de la psicología cognitiva (Boring, 2004).
- d. El uso cada vez más difundido de tecnologías de computación (incluyendo aplicaciones *online*) para la presentación de estímulos, obtención de respuestas y procesamiento automático de puntajes (Boring, 2004; Sánchez, s.f.).
- e. La aplicación de análisis estadísticos cada vez más potentes (análisis de reducción factorial, métodos descriptivos multivariados, regresiones log-lineales, modelos de ecuaciones estructurales, etc.), incluso si no se han resuelto los temas de fondo sobre cumplimiento de los niveles de medición requeridos (Sánchez, s.f.; Cañadas & Sánchez, 1998; Dagostín, 2005; Norman, 2010).
- f. Un uso más bien “informal” de ítems “similares a escalas” (tipo Likert, de diferencial semántico, etc.) en áreas de investigación aplicada, como

opinión pública o investigación de mercado, con un análisis más bien descriptivo y deslastrado del denso trasfondo teórico correspondiente (Dagostín, 2005).

REFERENCIAS

- Atorresi, H., Lozia, G., Abal, F., Galibert, M. & Aguerri, M. (2009). Teoría de respuesta al ítem: conceptos básicos y aplicaciones para la medición de constructos psicológicos. *Revista Argentina de Clínica Psicológica*, 18, 179-188.
- Boring, R.L. (2004). Cognition and psychological scaling: Model, method, and application of constrained scaling (Tesis doctoral). Carleton University, Ottawa, Canadá.
- Brenan, J. (1999). Historia y sistemas de la psicología. Recuperado de http://books.google.com.pe/books/about/Historia_y_sistemas_de_la_psicolog%C3%ADa.html?id=Q4CEsW3yiU8C
- Buela, G. & Sierra, J.C. (1997). Manual de Evaluación Psicológica: fundamentos, técnicas y aplicaciones. Recuperado de <http://books.google.co.ve/books?id=JfaKI4a5xBgC&pg=PA199&dq=Thurstone+Likert#PPA198,M1>
- Cañadas, I. & Sánchez, A. (1998). Categorías de respuesta en escalas tipo Likert. *Psichothema*, 10(3), 623-631. Recuperado de <http://156.35.33.98/reunido/index.php/PST/article/viewFile/7489/7353>
- Carifio, J. & Perla, R.J. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidots. *Journal of Social Sciences*, 3(3), 106-116.
- Chang, L. (1994). A psychometric evaluation of 4-point and 6-point Likert-type scales in relation to reliability and validity. [Abstract]. *Applied Psychological Measurement*, 18(3), 205-215. Recuperado de <http://apm.sagepub.com/content/18/3/205.short>
- Cook, S.W. & Selltiz, C. (1976). Enfoque de indicadores múltiples en la medición de actitudes. En G. Summers (Comp.). *Medición de actitudes*. Ciudad de México, México: Trillas.
- Cummins, R.A. & Gullone, E. (2000). Why we should not use 5-point Likert scales: The case for subjective quality of life measurement. Trabajo presentado en *Proceedings, Second International Conference on Quality of Life in Cities*, Universidad Nacional de Singapur, Singapur. Recuperado de http://www.deakin.edu.au/research/acqol/iwbg/correspondence/2002/Cummins_Gullone_2000_Likert_Scales.doc

- Dagostín, J. (2005). Influência do desempenho de atributos de produtos ou serviços na satisfação dos clientes: uma análise comparativa entre diferentes técnicas de pesquisa (Tesis de Maestría). Universidad Regional de Blumenau, Blumenau, Brasil. Recuperado de http://proxy.furb.br/tede/tede_busca/arquivo.php?codArquivo=220
- Dotson, L.E. & Summers, G.D. (1976). Cómo elaborar escalas técnicas de Guttman. En G. Summers (Comp.). *Medición de actitudes*. Ciudad de México, México: Trillas.
- Elejabarrieta, F.J. & Iñiguez, L. (1984). Construcción de escalas de actitud tipo Thurstone y Likert. Recuperado de <http://aprendeenlinea.udea.edu.co/revistas/index.php/ceo/article/viewFile/6820/6246>
- Fenton, N.E. & Pfleeger, S.L. (1997). Software metrics: A rigorous and practical approach. Londres, Inglaterra: PWS Pub. En Fernández, L. (1998). *Teoría de la medición*. Recuperado de <http://www.sc.edu.es/jjwdocoj/remis/docs/teoriamedicion.html>
- Garland, R. (1991). The mid-point on a rating scale: Is it desirable? *Marketing Bulletin*, 2, 66-70. Recuperado de http://marketing-bulletin.massey.ac.nz/V2/MB_V2_N3_Garland.pdf
- Göb, R., McCollin, C., & Ramalhoto, M.F. (2007). Ordinal methodology in the analysis of Likert scales. [Abstract]. *Quality & Quantity*, 41(5), 601-626. Recuperado de <http://link.springer.com/article/10.1007/s11135-007-9089-z#>
- González, V. (s.f.). Introducción a los métodos de escalamiento. Recuperado de <http://www.uv.es/gonzalev/PSICOM-04-05/PSICOM-TEMA%2003-INTRO-ESCAL.rtf>
- Guttman, L. (1976). Base para elaborar escalas con datos cualitativos. En G. Summers (Comp.). *Medición de actitudes*. Ciudad de México, México: Trillas.
- Heise, D. (1976). El diferencial semántico y la investigación de actitudes. En G. Summers (Comp.). *Medición de actitudes*. Ciudad de México, México: Trillas.
- Hernández, A., Espejo, B., González, R. & Gómez, J. (2001). Escalas de respuesta tipo Likert: ¿Es relevante la alternativa "indiferente"? *Metodología de Encuestas*, 3(2), 135-150. Recuperado de <http://casus.usal.es/pkp/index.php/MdE/article/viewFile/901/842>
- Jacoby, J. & Mattel, M. (1971). Three point Likert scales are good enough. [Abstract]. *Journal of Marketing Research*, 8(4), 495. Recuperado de <http://connection.ebscohost.com/c/articles/4997896/three-point-likert-scales-are-good-enough>

- Knapp, T.R. (1990). Treating ordinal scales as interval scales: An attempt to resolve the controversy. *Nursing Research*, 39, 121–123. Recuperado de http://www.mat.ufrgs.br/~viali/estadistica/mat2282/material/textos/treating_ordinal_scales%5B1%5D.pdf
- Muñiz, J. (1998). La medición de lo psicológico. *Psicothema*, 10(1), 1-21.
- Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. [Abstract]. *Advances in Health Sciences Education*, 15(5), 625-632. Recuperado de <http://link.springer.com/article/10.1007/s10459-010-9222-y>
- Osgood, C., Suci, G. & Tannenbaum, P. (1976). Medición de actitudes. En G. Summers (Comp.). *Medición de actitudes*. Ciudad de México, México: Trillas.
- Reyes, L. (2007). La teoría de acción razonada: implicaciones para el estudio de actitudes. *Investigación Educativa*, 7, 66-77. Recuperado de http://www.alfaguia.org/alfaguia/files/1320437914_40.pdf
- Sáiz, M. (2009). Historia de la psicología. Recuperado de http://books.google.co.ve/books/about/Historia_de_la_psicolog%C3%ADa.html?id=fZftd1tRrDMC
- Sánchez, M. (s.f.). Introducción a la teoría de respuesta al ítem, una herramienta para el análisis de variables latentes: aplicación a la medición de la calidad de vida de la infancia. Recuperado de <http://www.asepelt.org/ficheros/File/Anales/2004%20-%20Leon/comunicaciones/S%E1nchez%20Rivero%20Texto.pdf>
- Santoro, E. (1975). *El diferencial semántico*. Caracas, Venezuela: División de Publicaciones, Facultad de Ciencias Económicas y Sociales, Universidad Central de Venezuela.
- Thurstone, L.L. (1928). Attitudes can be measured. *American Journal of Sociology*, 33, 529-554. Recuperado de http://www.brocku.ca/MeadProject/Thurstone/Thurstone_1928a.html
- Wright, B. (s.f.). A history of social science measurement. Rasch measurement forum. Recuperado de <http://www.rasch.org/memo62.htm>