

Método de los Intervalos Aparentemente Iguales de Thurstone.

1. Introducción.

El método de los intervalos aparentemente iguales es un procedimiento de construcción de escalas para la medición de las actitudes que fue elaborado por Thurstone y presentado en diversos trabajos, pero especialmente en un importante artículo del año 1.928 titulado "Las actitudes pueden medirse". Aquí vamos a describir ordenadamente los pasos de este método de Thurstone que permite construir instrumentos para medir actitudes escalando los items para después medir los sujetos.

2. Generación de opiniones.

El primer paso consiste, desde luego, en *escoger y definir* aquella actitud sobre la que se desea elaborar una escala de medida.

Por supuesto, en tanto que se trata de una dimensión latente, la actitud no podrá ser medida directamente. Dicho de otro modo, las actitudes sólo pueden medirse indirectamente, a través de las opiniones, que, bajo ciertas circunstancias, pueden considerarse su manifestación verbal. Menos

frecuentemente las actitudes pueden medirse también a través de conductas que puedan considerarse su manifestación comportamental.

Para *generar opiniones* distintas que sirvan a la construcción de la escala, se comienza por seleccionar varios grupos, presumiblemente distintos en cuanto a la actitud objeto del estudio. Después se solicita a esta muestra (Muestra A) que haga una descripción escrita libre de sus opiniones hacia el objeto de estudio.

3. Construcción de items.

A partir del material escrito se seleccionan y elaboran un conjunto de 100 a 150 enunciados (Versión I) que expresen todas las posiciones posibles hacia el tema bajo consideración.

La elaboración de estos enunciados debe atenerse a determinadas ciertas recomendaciones. Después de haber confeccionado de 100 a 150 items se selecciona un número menor, entre 80 y 100 (Versión II), aplicando las recomendaciones acerca de las propiedades de los items de escalas de actitud.

A continuación se escribe cada enunciado en una tarjeta separada.

3.1. Propiedades para los items de actitud.

Desde el principio de la medición de actitudes muchos autores han formulado alguna lista de recomendaciones acerca de las cualidades y propiedades que deben tener los enunciados o "items" de una escala de actitud. La mayoría de esas listas tienen su origen en las recomendaciones iniciales que hizo el mismo Thurstone (1.928) y que a continuación exponemos sintetizadamente:

1. Los items deben cubrir el continuo completo de la actitud.

2. Hay que disponer de enunciados neutros que puedan ubicarse en el centro del continuo y evitar que la escala se rompa en dos según su signo.
3. Los enunciados deben ser breves para evitar la fatiga.
4. Han de construirse de tal modo que la respuesta de aceptación o rechazo dependa de la actitud del lector, de su aceptación o rechazo del objeto de actitud. Los items deben suscitar esa aceptación o rechazo.
5. La aceptación o rechazo del ítem no debe ser una cuestión de conocimientos o de información.
6. Hay que evitar los items ambiguos o de doble argumento.
7. Hay que evitar items ajenos al objeto de actitud bajo consideración.

En el capítulo 3 de la primera parte nos hemos ocupado más extensamente de estas cuestiones.

4. Clasificación de los items.

Se selecciona una muestra de 200 o 300 sujetos (Muestra B). Se solicita a cada uno de estos sujetos que ordene las tarjetas en once pilas ordenadas desde los enunciados más intensamente negativos hacia el objeto de actitud hasta los enunciados más intensamente favorables.

En palabras de Thurstone (1.928) "la tarea consiste fundamentalmente en distribuir las tarjetas en once pilas de tal forma que *parezcan* estar regularmente espaciadas o graduadas" Se etiquetan solamente la pila central (la número 6), como neutral, y las de los dos extremos, positivo y negativo.

Cada sujeto clasifica cada ítem en uno de los once montones graduados de 1 a 11.

Obsérvese que clasificar un ítem en una pila determinada es lo mismo que conferirle una puntuación a ese ítem entre 1 y 11. Por ejemplo, si el ítem A es clasificado por el sujeto p en el montón 5, puede decirse que el sujeto p dió al ítem A una puntuación de 5.

Es muy importante destacar que cada sujeto de la Muestra B no clasifica los ítems de acuerdo con su opinión personal sobre la cuestión, sino de acuerdo con el lugar que él cree que el ítem ocupa en el continuo, independientemente de que él personalmente esté o no de acuerdo con el ítem. No se trata de ubicar a los sujetos de la Muestra B respecto al continuo de actitud, se trata de que los sujetos de la muestra B ubiquen los ítems sobre un continuo de actitud de once puntos de extensión. Los miembros de la muestra B actúan como jueces de los ítems, tan independientemente como sea posible de su propia ubicación personal en el continuo de favorabilidad o desfavorabilidad hacia el objeto de actitud (Tarea de Juicio).

Si se consideran las once pilas como once alternativas graduadas, -aunque sin etiquetar verbalmente excepto la central y los extremos-, entonces puede afirmarse que el método de los intervalos aparentemente iguales de Thurstone utiliza una *tarea de elección de alternativas con orientación de juicio* (Método II.A) como método de obtención de información para escalar los ítems. Sin embargo, en la concepción de Thurstone el sujeto es quien establece la distancia entre las "alternativas"; este matiz puede ser importante porque implica cómo se concibe la tarea psicológica que efectúa el sujeto. Sin embargo, el tratamiento que da a la información, a mi juicio, olvida ese matiz teórico.

5. Calibración de los ítems.

Para cada uno de los ítems disponemos de las puntuaciones que los sujetos les han conferido. Por ejemplo, con una Muestra B de 200 sujetos disponemos de 200 puntuaciones para cada ítem.

Podemos estudiar ahora la distribución de puntuaciones que la Muestra B ha dado a cada ítem. Si para un ítem determinado ubicamos en abscisas la escala de 1 a 11, y en ordenadas la frecuencia de cada puntuación, el número de sujetos que conceden cada una de las once posibles puntuaciones, podremos trazar el polígono de frecuencias de cada uno de los ítems. Esto es, podremos representar gráficamente la distribución de puntuaciones que recibe cada ítem.

Las puntuaciones que recibe cada ítem por la Muestra B se utilizarán para caracterizar al ítem. Si un ítem es "muy negativo" habrá sido clasificado como muy negativo por la mayoría de los sujetos de la Muestra B, y, por tanto, habrá recibido puntuaciones cercanas a 1. Si un ítem es neutro habrá recibido la mayoría de sus puntuaciones en torno a 6, que es la categoría neutral. Si un ítem es "muy positivo" respecto al objeto de actitud tendrá puntuaciones altas cercanas a 11.

Por supuesto, definir el 1 como el extremo negativo y el 11 como el positivo es arbitrario, con la única condición obvia de que esa definición permanezca constante a lo largo del estudio.

Cada ítem tiene, pues, una distribución de puntuaciones. Cada una de esas distribuciones describe la ubicación de cada ítem en el continuo de actitud. Obviamente, en la distribución correspondiente a cada ítem se pueden calcular los estadísticos que resulte oportuno.

Los estadísticos de posición central indicarán la ubicación del ítem y los de dispersión el grado de acuerdo o desacuerdo que hubo entre los miembros de la Muestra B.

Thurstone toma la *mediana* como valor de escala del ítem. Es decir, para un ítem, por ejemplo, que sea ubicado por el 50% de los sujetos de la Muestra B por debajo del valor 7'33, -quedando necesariamente el 50% restante de las puntuaciones por arriba de 7'33-, tomamos como su valor de escala 7'33. El ítem está "ubicado", en término medio, en 7'33.

De este modo tenemos descrita por su mediana la posición de cada ítem en el continuo de 1 a 11. Tenemos todos los ítems calibrados, colocados en un

lugar preciso de la dimensión de actitud. La calibración de cada ítem indica exactamente cuán positivo o negativo es hacia el objeto de actitud, de acuerdo con la determinación empírica obtenida mediante las puntuaciones de la Muestra B.

Por supuesto es posible utilizar la *media* de la distribución de puntuaciones de un ítem como valor de escala del mismo. Si la distribución de puntuaciones dadas a un ítem es normal, como Thurstone esperaba, entonces ambas, media y mediana, coincidirán.

Es posible poner reparos al uso de la media como estadístico en este proceso de construcción de una escala debido a que sólo el juicio subjetivo de los sujetos garantiza la constancia de la distancia entre cada dos valores de las once pilas. En la medida en que no pueda garantizarse una distancia, una unidad de medida, estable y constante entre 1 y 2, entre 2 y 3, etc., según la teoría de los estadísticos admisibles de Stevens, no sería correcto calcular la media. En la medida en que los intervalos entre los once valores sean constantes, el uso de la media es correcto y presenta las ventajas propias de este estadístico.

Cuanto más simétrica sea la distribución, sea o no normal, tanto más coincidirán media y mediana. Si la distribución no es simétrica el valor de escala para un ítem variará en función de cual de ambos estadísticos elijamos. Deben considerarse entonces, para su elección, las propiedades características de estos estadísticos, de las que no corresponde tratar aquí. Sin embargo, puede recordarse que, principalmente, la mediana aportará una mayor robustez a la determinación del valor de escala, mientras que la media aportará una mayor eficiencia.

Este proceso que permite ubicar a cada ítem en algún punto de la recta que representa el continuo de actitud es lo que se denomina calibrar o "escalar" un ítem.

6. Evaluación de la constancia de las calibraciones.

Thurstone era consciente de la dificultad que suponía que la Muestra B actuase independientemente de sus propias convicciones acerca del objeto de actitud. Es decir ¿estarían las calibraciones de los items influidas de un modo importante por las propias actitudes de los sujetos de la Muestra B?

Por ejemplo, en una escala de actitudes hacia la religión ¿quedarían influidas las calibraciones de los items porque la muestra estuviese formada por religiosos o por ateos? ¿clasificarían los sujetos de ambos grupos del mismo modo los items independientemente de que estuvieran o no de acuerdo con ellos?

Thurstone sugirió que esto debía ponerse a prueba empíricamente. Para ello, los items deberían escalarse de modo independiente con dos muestras representativas de los dos extremos de actitud. Si los items quedaban igualmente calibrados por ambas muestras, entonces podría admitirse la calibración, en caso contrario la calibración sería dependiente de la muestra que actúa como juez.

En el ejemplo anterior, las calibraciones derivadas de una Muestra B de ateos y de otra Muestra B de religiosos, realizadas separadamente, deberían permanecer suficientemente próximas para que pudiésemos aceptar su generalidad.

En la práctica puede no ser necesario contar propiamente con dos Muestras B. Bastará con separar en una muestra B los sujetos que -por alguna razón o información ajena a la escala- se sepa que pertenecen a grupos con actitudes en los extremos. Si tenemos el suficiente número de personas de ambos grupos es posible calcular la calibración de los items independientemente en ambos colectivos. Aquellos items que mantienen las calibraciones constantes podría admitirse que han sido calibrados con independencia de las actitudes personales de los sujetos de la muestra B. Aquellos con calibraciones discrepantes no podríamos incluirlos en la escala definitiva, debido a que no podríamos afirmar con seguridad cuál es su valor escalar.

Un procedimiento como el de Thurstone garantiza tanto más la calibración escalar de un ítem cuanto más estable haya permanecido su valor de escala a través de más muestras B de calibración. En este aspecto, la constancia de las calibraciones en diversas muestras aumenta nuestra confianza en la escala.

En general, los valores escalares obtenidos por cualquier método de escalamiento nos resultan más confiables si el estudio ha sido replicado a través de diversas muestras y a través de diversos métodos.

7. Evaluación de la ambigüedad de los ítems.

No sólo se trata de que un ítem tenga una calibración estable a través de muestras. También importa que la calibración haya permanecido lo más estable posible para los diferentes sujetos de una misma muestra.

Si un ítem es, por ejemplo, claramente neutro, todos los sujetos tenderán a situarlo en la pila 6, a ubicarlo como neutro y a concederle "seis puntos" en la escala. En el caso extremo de un ítem cuya interpretación como neutro fuera inequívoca y constante para todos los sujetos, todos le concederían la puntuación de 6. Por supuesto, la mediana de la distribución sería 6 y el valor escalar del ítem, su ubicación en la escala, sería 6. La dispersión de la distribución sería nula. La unanimidad o el acuerdo de los jueces parece garantizar que, efectivamente, el valor de escala del ítem está bien puesto.

Imaginemos el caso opuesto. Si un ítem es tan ambiguo que no existe ningún acuerdo entre los sujetos respecto a su clasificación, podemos esperar que aparezcan puntuaciones del ítem a lo largo de toda la escala, desde el uno hasta el once. En el caso ideal de un ítem completamente ambiguo, la ubicación dada por los sujetos podría calificarse desde un observador externo como emitida al azar; y la probabilidad de que un sujeto ubicase al ítem en cada una de las once pilas sería la misma, es decir, $1/11 = 0'09$. Es decir, se produciría una distribución plana para las calibraciones de ese ítem; de cada 100 sujetos aproximadamente 9 lo ubicarían en cada uno de los once niveles. La mediana del ítem también estaría idealmente en 6, como en el

caso anterior, pero ahora el ítem presentaría una distribución con una fuerte dispersión.

Es obvio que el primer ítem es preferible al segundo. De hecho el segundo ítem no serviría para nada debido a su total ambigüedad. En síntesis, puede desprenderse de estos ejemplos que cuanto menor sea la dispersión de la distribución de puntuaciones de cada ítem tanto mejor. Cuanto mayor es esa dispersión más ambiguo es el ítem, menos claro está qué significa en términos de actitud, y, por tanto, resulta menos adecuado.

Esto quiere decir que además de los estadísticos de tendencia central, como la mediana, que indican el valor escalar del ítem, también debe atenderse a los estadísticos de dispersión, para descartar los ítems con mucha dispersión.

Se puede utilizar el rango semi-intercuartil y/o la desviación típica para considerar esta dispersión.

El rango semi-intercuartil es una medida de dispersión que establece cual es la mitad de la distancia escalar entre los cuartiles uno y tres. Es el estadístico de dispersión paralelo a la mediana como estadístico de posición central en la construcción de una escala de Thurstone.

La desviación típica es el estadístico de dispersión correspondiente al uso de la media como estadístico de posición central para determinar el valor de escala. Aquí pueden repetirse respecto de la desviación típica las consideraciones que hemos hecho anteriormente respecto al uso de la media.

En resumen, se estudia la ambigüedad de los ítems para eliminar los ítems ambiguos, es decir, los de mayor dispersión, dejando los ítems cuya valoración es más inequívoca.

8. Evaluación de la irrelevancia de los ítems.

Un ítem puede presentar un determinado valor de escala, una reducida dispersión y todavía no ser adecuado debido a su irrelevancia. La

irrelevancia es la cualidad que hace a ciertos ítems no ser escogidos en una tarea de respuesta, comparativamente con otros de su mismo valor de escala.

Para evaluar la irrelevancia de los ítems de un modo objetivo Thurstone sugirió el siguiente procedimiento. Se toma la colección de unos 100 ítems bajo estudio (Versión II), y se administra a una muestra (Muestra C) de "varios centenares" de sujetos, para que digan si están de acuerdo o no con cada ítem (Tarea de Respuesta).

Esta Muestra C puede ser o no, Thurstone admite ambas posibilidades, la misma que la muestra B. Sin embargo, la tarea que se pide ahora es completamente distinta. Ahora se solicita a cada sujeto que diga si está de acuerdo o no con cada ítem. Esta operación es la propia que se utilizará después para medir a los sujetos, para ubicar a los sujetos en el continuo de actitud. Thurstone en concreto solía pedir a sus sujetos que marcaran con un "+" aquellos ítems que estaban de acuerdo con sus ideas y con un "-" aquellos otros con los que estaban en desacuerdo.

Cada ítem tenía un valor de escala que fue obtenido en el proceso de calibración. Pues bien, la versión más simple de la idea central de irrelevancia puede expresarse del siguiente modo. *Si dos ítems con la misma calibración son igualmente relevantes entonces una misma proporción de sujetos de una Muestra C estará de acuerdo con ellos.*

Si dos ítems tienen un mismo valor de escala bien determinado pero los sujetos no admiten uno de ellos en proporción semejante al otro, el ítem menos aceptado es posible que sea "irrelevante", quizás afectado por aspectos distintos al que se pretende medir.

Esta sencilla idea fue desarrollada por Thurstone de un modo un poco más complejo. Esta segunda aproximación a la evaluación de la irrelevancia tiene en cuenta la distribución de "elecciones" que presentan los diversos ítems ya graduados en una tarea de respuesta.

Una vez administrada la escala (Versión II) ya calibrada a una muestra C, para cada ítem habrá una frecuencia, y por tanto una proporción, de sujetos que "admitan el ítem", de sujetos que muestran su acuerdo con el mismo

como representación de su actitud. Tomemos por ejemplo un ítem cuyo valor escalar es 5. Supongamos una muestra C de 200 sujetos y supongamos que 100 de ellos estuvieron de acuerdo con el ítem en cuestión. Si consideramos sólo la parte de la muestra que está de acuerdo con ese ítem, los cien sujetos que dijeron que sí al ítem, entonces, la proporción de esos sujetos que esté de acuerdo con otro ítem cualquiera será una función continua inversa de la distancia del valor escalar de ese otro ítem al valor escalar del ítem tomado como referencia.

En términos de ejemplo, es consistente que de los cien sujetos que dijeron sí al ítem de valor escalar 5, ochenta -por ejemplo- digan sí a un ítem de valor 4, setenta -por ejemplo- a un ítem de valor 3, cuarenta a un ítem de valor 2 y diez a un ítem de valor 1. Algo semejante debería suceder también en el tramo superior de valores escalares a partir de cinco. Es decir, a más distancia en valor escalar menos sujetos se comportarán del mismo modo. Las desviaciones severas de este patrón de aceptación para algún ítem pueden indicar problemas de irrelevancia.

El concepto de irrelevancia es utilizado por Thurstone de un modo amplio para identificar el comportamiento inusual de un ítem para su ubicación y dispersión escalar.

9. Selección de ítems y elaboración de la escala de medida de actitudes.

En los pasos anteriores hemos obtenido un conjunto de valiosas informaciones sobre los aproximadamente cien ítems (Versión II) que seleccionamos al principio a partir de una muestra inicial de ítems mucho mayor (Versión I). Ahora se trata de usar las informaciones obtenidas en los pasos anteriores para seleccionar un conjunto más reducido de veinte o treinta ítems (Versión III) que formen la escala de medida propiamente dicha. Para ello:

Primero, se eliminan los ítems con demasiada ambigüedad.

Segundo, se eliminan los items que hayan mostrado ser "irrelevantes" para la escala.

Tercero, de los items restantes, todos ellos relevantes y no ambiguos, se seleccionan entre veinte y treinta, de tal forma que sus calibraciones estén repartidas de forma tan uniforme y equidistante como sea posible a lo largo de todo el continuo de la escala.

En este punto se dispone ya de una escala destinada a la medición de actitudes con todos sus items calibrados y bien distribuidos a lo largo del continuo de actitud.

10. Medición de actitudes.

La escala (Versión III) obtenida como fruto del paso anterior está ya lista para medir las actitudes de una muestra de sujetos (Muestra D).

Para medir sujetos, simplemente se administra la lista de items a los sujetos solicitándoles que indiquen con qué items están de acuerdo y con qué items no. Qué items representan su propia actitud. Un sujeto debe decidir ante cada ítem si éste refleja o no su punto de vista. Los items con los que el sujeto manifiesta estar de acuerdo se entiende que reflejan la actitud del sujeto (*Tarea de estímulo simple con orientación de respuesta, I.B*).

Para ubicar a cada sujeto, "medirlo" o "escalarlo", en el continuo de actitud necesitamos aprovechar la información de calibración de los items. Decir que sí a un ítem cuyo valor de escala es 6'7 significa estar en el valor 6'7 de la escala. Extendiendo esta idea, se toma como puntuación o ubicación de la persona en el continuo de actitud la media aritmética de los valores de escala de aquellos items con los que ha manifestado acuerdo. Esa puntuación sitúa al sujeto en la dimensión de actitud del mismo modo que la calibración había situado a los items (*Procedimiento H2; I:1; II:2; III:2*).

Obsérvese que el proceso de construir la escala, pasos uno al nueve, y el uso de medición de la misma, paso diez, son cosas bien distintas. Es posible aplicar la escala (Versión III) sobre las muestras A, B y C que antes mencionamos para medir a los sujetos de esas muestras, pero desde luego la escala esta lista ahora para medir otras muestras distintas de aquellas que participaron en esta costosa elaboración del instrumento de medida.

Por supuesto, una vez elaborada la escala de medida (Versión III) con ella se pueden medir grupos distintos, obtener distribuciones de puntuaciones para esos grupos y compararlos en sus actitudes. Se pueden efectuar obviamente los tipos de estudios propios de la medición de una característica psicológica.

11. Algunas observaciones adicionales al proceso de construcción de la escala.

En los pasos de trabajo anteriores, que fueron propuestos por Thurstone, deberían añadirse tres aspectos importantes más que actualmente no dejarían de ser considerados por un investigador. Esos tres aspectos se refieren a la evaluación de la dimensionalidad de la escala y a la obtención de pruebas acerca de su fiabilidad, y, muy especialmente, de su validez.

La *cuestión de la dimensionalidad* es compleja y ha suscitado una inmensa cantidad de investigación. La cuestión es aproximadamente, como ya hemos señalado, si el conjunto de items elaborado por el procedimiento anterior necesariamente está correctamente representado en una única dimensión.

En la actualidad resultaría usual someter a las puntuaciones de una o más muestras D a análisis factorial, en sentido amplio, o a técnicas de escalamiento multidimensional, con el propósito de estudiar si representaciones no unidimensionales resultan más adecuadas.

Este tipo de información puede contribuir a excluir lo que Thurstone trataba bajo el no muy transparente término de "irrelevancia" y a descubrir aspectos del funcionamiento de los items que mejoren la escala.

La *cuestión de la fiabilidad* es bien conocida para quien ha seguido alguna vez un curso elemental de teoría clásica de tests. El estudio de la fiabilidad implica dos grandes campos: la cuestión de la consistencia interna del instrumento y la cuestión de la estabilidad temporal de las puntuaciones.

Thurstone sugirió usar un procedimiento de formas paralelas para tratar la fiabilidad de sus escalas. Obviamente, otros procedimientos para evaluar la fiabilidad orientados a la consistencia interna y orientados a la estabilidad temporal también pueden ser aplicados y estudiados.

En particular los procedimientos relativos a la estabilidad temporal, lo que se denomina a veces test-retest, deben ser objeto de un estudio cuidadoso en el caso de las escalas de actitudes por dos razones. Primero, porque las actitudes pueden ser sensibles a cambios situacionales más o menos importantes, y, en ese sentido, las evaluaciones para periodos relativamente largos pueden diferir porque los sujetos realmente cambian. Segundo, porque en periodos cortos, especialmente con escalas con pocos items, factores de memoria e intencionalidad pueden actuar sobre las segundas puntuaciones. No es que estos problemas sean exclusivos del campo de las actitudes, pero pueden sumarse a otros como la insinceridad, y dificultar el estudio de la fiabilidad.

La *cuestión de la validez* es, desde luego, muy importante. Thurstone trata la cuestión de la independencia de las calibraciones de la Muestra B, la cuestión de la ambigüedad de los items y la cuestión de la irrelevancia, como parte del estudio de la validez.

Además, en la actualidad, el estudio de la validez debería implicar estudios más amplios que consideraran al menos la relación de las puntuaciones de los sujetos (Muestra D) en la actitud bajo estudio con otras variables consideradas como criterios adecuados.

Diversos tipos de estudios de validez pueden proponerse para las escalas tipo Thurstone. Un tipo importante de estudio consiste en medir, para una Muestra D, además de la actitud bajo estudio, otras actitudes y comportamientos que se supone por hipótesis que están relacionados, procediendo a la puesta a prueba de esas hipótesis. Si el cuestionario no mantuviera las relaciones esperadas con otras medidas consideradas

aceptables, realizadas mediante otras pruebas, difícilmente podría defenderse su validez. Otro tipo importante de estudio consiste en utilizar grupos extremos de los que, por algún medio externo bien establecido, se conoce su actitud. Aquí se trata de evaluar la relación entre las puntuaciones en la escala y las actitudes que por algún otro medio se sabe que sustentan esos grupos. Si una escala sobre un tema determinado es incapaz de discriminar adecuadamente a los miembros de grupos extremos en esa actitud se pondría en duda su capacidad de medir adecuadamente la actitud pretendida.

La cuestión de la validez es compleja y múltiple. Una escala tiene diferentes coeficientes de validez con diferentes criterios y un conjunto de métodos diferentes y de resultados no siempre acordes pueden ser usados para su estudio. Desde el trabajo de Cronbach y Meehl de 1.955 estas cuestiones suelen tratarse en el marco de la llamada validez de constructo.

Como ya pusimos de manifiesto la cuestión de la dimensionalidad, de la fiabilidad y de la validez forman parte del proceso de evaluación de la escala, del que no nos ocuparemos aquí, y afectan a la construcción de escalas utilizando cualquier método de escalamiento.

Además de las cuestiones generales de la dimensionalidad, la fiabilidad y la validez hay otras también importantes de las que ya nos ocupamos al tratar del método de las comparaciones apareadas y que pueden trasladarse aquí con pocos cambios. Entre ellas se cuentan la necesidad de un cumplimiento suficiente de los *supuestos del método*, las dificultades asociadas al uso de los *sujetos como jueces* y calibradores de los items, y los *costes de esfuerzo y tiempo* que la aplicación del método conlleva.

Cap. 10. METODO DE LOS INTERVALOS APARENTEMENTE IGUALES DE THURSTONE.

ELABORACION DE LA ESCALA:

Generación de ítems:

1. Generación de opiniones sobre la cuestión a medir (Muestra A).
2. Construcción de los ítems: de 100 a 150 (Versión I).
3. Selección racional de los ítems: de 80 a 100 (Versión II).

Obtención de información para escalar los ítems:

1. Clasificación de los ítems (Muestra B: de 200 a 300 sujetos).
"Distribuir los ítems en 11 pilas de tal forma que parezcan estar regularmente espaciadas o graduadas, desde muy desfavorable hasta muy favorable ante el objeto de actitud, con la pila intermedia considerada como neutra" (La tarea se puede concebir como una *elección de alternativas con orientación de Juicio: II.A*).

Obtención de los valores de escala de los ítems:

1. Disposición de la información obtenida:
Se elabora una tabla "ítems por jueces" en la que cada celdilla contiene el valor (de 1 a 11) asignado por cada juez a cada ítem.
2. Calibración de los ítems:
La mediana de la distribución de puntuaciones que recibe cada ítem es su valor de escala (*b*). (También se puede utilizar la media).

Valoración y selección de ítems:

1. Evaluación de la ambigüedad de los ítems:
El rango semi-intercuartil de la distribución de puntuaciones que recibe cada ítem evalúa la ambigüedad del ítem (También se puede utilizar la desviación típica).
Se prefieren los ítems poco ambiguos, con poca dispersión.
2. Evaluación de la constancia de las calibraciones:
Se realiza el proceso de calibración en grupos extremos separadamente.
Se prefieren los ítems que mantienen constantes sus calibraciones.
3. Evaluación de la irrelevancia de los ítems:
Dos ítems con la misma calibración son igualmente relevantes si presentan el mismo número de elecciones en una muestra C. Se prefieren los ítems relevantes, es decir, que son utilizados por los sujetos de una muestra C.
4. Selección de los ítems (Versión III):
Utilizando los tres criterios anteriores y, además, procurando disponer de ítems graduados a intervalos regulares a través de todo el continuo.

MEDICION DE SUJETOS:

Obtención de información:

1. Se administra la escala (Versión III) a los sujetos a medir (Muestra D) y se les solicita que expresen si cada ítem los representa o no. (P.e., si reflejan o no la actitud del sujeto).
(*Tarea de estímulo simple con orientación de Respuesta: I.B*).

Obtención de los valores de escala o puntuaciones de los sujetos:

1. Disposición de la información obtenida:
Se elabora una tabla "ítems por sujetos" en la que cada celdilla contiene el valor (de 0 ó 1) que representa la respuesta (acuerdo o desacuerdo) de cada sujeto con cada ítem.
2. Puntuaciones de los sujetos:
El promedio de los valores de escala de los ítems que representan a un sujeto se consideran la puntuación del sujeto (su ubicación sobre la dimensión medida).
(*Método H2; I:1; II:2; III:2*)