

TEMA 10. MÉTODOS DE ESCALAMIENTO DE SUJETOS

(Pasos a seguir para la construcción de un test)

10.1. Definición del constructo que se pretende medir

10.2. Establecer la finalidad para la que se construye el test

10.3. Fijar las posibles restricciones de la prueba

10.4. Establecer un completo plan detallado

10.5. Decidir el formato de respuestas de los ítems

10.6. Elaborar un número suficientemente amplio de ítems

10.7. Analizar empíricamente los ítems elaborados

10.8. Seleccionar los ítems definitivos

10. Pasos a seguir para la construcción del test:

10.1. Definición del constructo que se pretende medir

10.2. Establecer la finalidad para la que se construye el test

- Instrucción
- Diagnóstico
- Selección
- Colocación
- Clasificación
- Orientación
- Investigación
- etc.

10.3. Fijar las posibles restricciones de la prueba

- Tiempo limitado o ilimitado
- Poblaciones específicas o generales
- Aplicación individual o colectiva
- etc.

10.4. Establecer un completo plan detallado.

- Especificación de contenidos
- Especificación de procesos
- etc.

10. Pasos a seguir para la construcción del test (cont.)

10.5. Decidir el formato de respuestas de los ítems

- Respuesta libre (producción).
- Respuesta múltiple (elección).
- Respuesta dicotómica
- Escala tipo Likert
- Escalas graduadas
- etc.

10.6. Elaborar un número suficientemente amplio de ítems

Teniendo en cuenta que una vez que sometamos a análisis empíricos los ítems elaborados nos veremos obligados a eliminar aquellos que no alcancen unos requisitos mínimos de calidad psicométrica, se trata de elaborar un número suficientemente amplio de ítems que garantice que dispondremos de un repertorio con calidad suficiente para establecer el test definitivo. Para ello se recomienda elaborar el doble o triple de ítems de los que previamente se hayan considerado para constituir el test definitivo.

10. Pasos a seguir para la construcción del test (cont.):

10.7. *Analizar empíricamente los ítems elaborados*

- selección de una muestra
- aplicación de los ítems
- obtención de las puntuaciones:

Puntuación de cada ítem:

valor asignado a las respuestas de los sujetos en función del formato de los ítems

Puntuación Total:

En general la puntuación total será la suma de las puntuaciones obtenidas por el sujeto en cada ítem.

En ítems de ejecución máxima y con formato de elección múltiple se suele eliminar el efecto del acierto por azar. En este caso, la puntuación total será:

$$X_T = A - \frac{E}{k-1}$$

En donde X_T es la puntuación total

A es el número de ítems acertados

E es el número de ítems fallados

k es el número de opciones de respuesta de los ítems

10. Pasos a seguir para la construcción del test (cont.):

10.7. Analizar empíricamente los ítems elaborados (cont.)

- análisis de la calidad psicométrica de los ítems (tres tipos de indicadores):

10.7.1. Índices que describen la distribución de las respuestas a un elemento único: Poder discriminativo e índice de dificultad (este último sólo para ítems dicotómicos de ejecución máxima).

10.7.2. Índices que describen el grado de relación entre las respuestas al ítem y algún criterio de interés: Índice de discriminación interna (cuando el criterio es la puntuación total del resto de ítems) e índice de discriminación externa (cuando el criterio es una medida externa)

10.7.3. Índices combinados de los anteriores (índices de fiabilidad y validez de los ítems)

(Veamos cada uno de ellos con cierto detenimiento)

10. Pasos a seguir para la construcción del test (cont.):

10.7. Analizar empíricamente los ítems elaborados (cont.)

10.7.1. Índices que describen la distribución de las respuestas a un elemento único (cont.)

- Poder Discriminativo del ítem (PD_i): Es la variabilidad de las puntuaciones asignadas a ese ítem. Es decir es la Desviación Típica del ítem (S_i).

En distribuciones binomiales (dicotómicas), la Desviación Típica es:

$$PD_i = S_i = \sqrt{pq}$$

Donde p es la proporción de 1 y q es la proporción de 0 ó, lo que es lo mismo, el complementario de p , ($q = 1 - p$).

El PD_i se interpreta en función de la Variabilidad máxima ($S_{i\max}$) que puede tomar un ítem y que depende del nº de valores que se puedan asignar a dicho ítem. Ambos siguen una progresión aritmética; partiendo del mínimo que son dos posibles valores y cuya $S_{i\max}$ es 0,5, por cada unidad que aumentan los valores, aumenta media unidad la Variabilidad Máxima:

Nº de valores:	2	3	4	5	6	7	8	etc.
$S_{i\max}$:	0.5	1	1.5	2	2.5	3	3.5	etc.

El poder discriminativo de un ítem suele expresarse en porcentajes (%). Basta para ello con dividir la S_i por su máximo ($S_{i\max}$) y multiplicar el resultado por 100.

Ejemplo

Un ítem “a” con 5 opciones de respuesta de la que sólo una es correcta tiene una desviación típica $S_i = 0,35$.

Otro ítem “b” con formato tipo Likert con 5 categorías de respuesta tiene una desviación típica $S_i = 1,2$.

¿Cuál de los dos ítems tiene mayor poder discriminativo?

Solución:

$$PD_i = \frac{S_i}{S_{i\max}} 100$$

El ítem a tiene un poder discriminativo de 0.35, siendo su límite máximo 0.50 ya que las respuestas de los sujetos sólo se valoran de forma dicotómica: Acierto = 1; Error = 0

El ítem b, sin embargo, tiene un poder discriminativo de 1.2, pero su límite máximo es 2 ya que las respuestas de los sujetos, al ser una escala tipo Likert, pueden recibir 5 valores distintos.

Por lo tanto:

$$PD \text{ ítem a} = (0,35/0,5) \cdot 100 = 70\%$$

$$PD \text{ ítem b} = (1,20/2,0) \cdot 100 = 60\%$$

Tiene mayor PD el ítem a

10. Pasos a seguir para la construcción del test (cont.):

10.7. Analizar empíricamente los ítems elaborados (cont.)

10.7.1. Índices que describen la distribución de las respuestas a un elemento único (cont.)

- Índice de Dificultad del ítem (ID_i): (Sólo aplicable a ítems dicotómicos de tests de ejecución máxima). Es la proporción de aciertos al ítem (p_A). Habría que hablar más propiamente de índice de facilidad que de dificultad ya que sus valores oscilan entre 0 y 1, tendiendo a 0 cuanto más difícil es el ítem y a 1 cuanto más fácil es.

$$ID_i = p_a = \frac{A}{N}$$

donde:

P_a es la proporción de aciertos

A es el nº de sujetos que aciertan el ítem

N es el nº de sujetos que responden al ítem

Un test equilibrado debería tener una dificultad intermedia, es decir, estar integrado por ítems de dificultad media, en torno a 0,5, porque son aquéllos que tienen mayor poder discriminativo. Sin embargo esta solución no es operativa por lo que conviene que el test esté integrado mayoritariamente por ítems con un rango de dificultad aceptable que cubra el espectro total.

En este sentido se recomienda que el 80% de los ítems de un test presenten de forma equilibrada dificultades que oscilen entre 0,25 y 0,75; reservando el 20% restante para introducir ítems de dificultades más extremas que permitan establecer un efecto “suelo” y un efecto “techo”.

10. Pasos a seguir para la construcción del test (cont.):

10.7. Analizar empíricamente los ítems elaborados (cont.)

10.7.2. Índices que describen el grado de relación entre las respuestas al ítem y algún criterio de interés (cont.)

- Índice de Discriminación Interna (IDI_i): Se le llama también Índice de Homogeneidad. El criterio de interés con el que se relaciona el ítem es la puntuación total del propio test al que pertenece el ítem.

El índice estadístico más utilizado es la correlación. Dependiendo del tipo de datos que analicemos se emplea cualquier coeficiente de correlación, aunque los más usados son la correlación producto-momento de Pearson (r), la correlación biserial (r_b) y la correlación biserial-puntual (r_{bp})

Interpretación:

Lógicamente cuanto mayores sean los índices de discriminación mejor. En cualquier caso, para que un ítem sea aceptable deberá tener un índice, como mínimo, estadísticamente significativo.

Ese valor mínimo depende del tamaño de muestra y del nivel de significación “alfa” (α) con el que queramos trabajar.

10. Pasos a seguir para la construcción del test (cont.):

10.7. Analizar empíricamente los ítems elaborados (cont.)

10.7.2. Índices que describen el grado de relación entre las respuestas al ítem y algún criterio de interés (cont.)

- Índice de Discriminación Interna (cont)

Los valores mínimos estadísticamente significativos son los siguientes:

Para la correlación de Pearson y para la biserial puntual (r_{ixmin} y $r_{bpixmin}$):

$$r_{bpixmin} = r_{ixmin} = z_{1-\frac{\alpha}{2}} \sqrt{\frac{1}{N-1}}$$

Para la correlación biserial (r_{bixmin}):

$$r_{bixmin} = \frac{0.5 r_{bpixmin}}{0,3989}$$

Ejemplo

Tenemos tres items (a, b y c) aplicados a 60 sujetos.

Sus índices de discriminación interna son, respectivamente, $r_{ax} = 0,28$; $rb_{bx} = 0,29$ y $rbp_{cx} = 0,25$

¿Qué podemos concluir acerca de su discriminación interna con $\alpha = 0,05$?

Solución: Para responder tenemos que averiguar el valor mínimo estadísticamente significativo. Si los valores de nuestros items superan esos valores mínimos serán aceptables.

$$r_{ix \min} = 1,96 \sqrt{\frac{1}{60-1}} = 0,26$$

$$rbp_{ixmin} = r_{ixmin} = 0,26$$

$$rb_{ix \min} = \frac{rbp_{ix \min} 0,5}{0,3989} = \frac{0,26 \times 0,5}{0,3989} = 0,32$$

Conclusión:

El item a es aceptable porque $0,28 > 0,26$

El item b es inaceptable porque $0,29 < 0,32$

El item c es inaceptable porque $0,25 < 0,26$

10. Pasos a seguir para la construcción del test (cont.):

10.7. Analizar empíricamente los ítems elaborados (cont.)

10.7.2. Índices que describen el grado de relación entre las respuestas al ítem y algún criterio de interés (cont.)

- Índice de Discriminación Externa (IDE_i): El criterio de interés con el que se relaciona el ítem es una medida ajena al test pero con la que se supone que el test está relacionado. Es como un coeficiente de validez del ítem (r_{iy})

El índice estadístico más utilizado es la correlación. Dependiendo del tipo de datos que analicemos se emplea cualquier coeficiente de correlación, aunque los más usados son la correlación producto-momento de Pearson (r), la correlación biserial (r_b) y la correlación biserial-puntual (r_{bp})

Interpretación: Se interpreta igual que el índice de homogeneidad o de discriminación interna

10. Pasos a seguir para la construcción del test (cont.):

10.7. *Analizar empíricamente los ítems elaborados (cont.)*

10.7.3. Índices combinados de los anteriores : índices de fiabilidad y validez de los ítems (cont.)

- Índice de Fiabilidad del ítem (IF_i): Nos informa de la contribución que ese ítem aporta a la medida del test total.

Es una combinación del Poder Discriminativo del ítem y su Discriminación Interna:

$$IF_i = S_i \cdot r_{ix}$$

- Índice de Validez del ítem (IV_i): Nos informa de la capacidad del ítem para contribuir a la predicción por parte del test del criterio externo de validación.

Es una combinación del Poder Discriminativo del ítem y su Discriminación Externa:

$$IV_i = S_i \cdot r_{iy}$$

Interpretación:

Ambos se interpretan exactamente igual que el Poder Discriminativo de los ítems, empleando además las mismas ecuaciones para su interpretación, ya que tanto el índice de fiabilidad máximo como el índice de validez máximo de cualquier ítem coinciden con S_{imax}

10. Pasos a seguir para la construcción del test (cont.):

10.8. Seleccionar los ítems definitivos

Una vez analizados todos los ítems elaborados, procedemos a seleccionar los definitivos que constituirán el test. Para ello se procede a elegir aquellos que mejor calidad psicométrica tengan en función de los análisis realizados.

Si disponemos de un número suficientemente grande de ítems que cumplan ese criterio de forma adecuada y no necesitamos tantos para construir el test definitivo, procederemos a elegir aquellos que desde un punto de vista cualitativo y de contenido más nos convengan.

Con el test ya elaborado, completamos el estudio psicométrico con el análisis de sus propiedades métricas como instrumento de medida único: ESTUDIO DE SU FIABILIDAD Y DE SU VALIDEZ.

Finalmente, si el instrumento de medida es bueno, procederemos a establecer, si procede, los baremos o normas de interpretación de las puntuaciones que nos reporta.