FF 17 08

Asignatura, Pedagogía Experimental. Título, Validez de los Tests. Guión número 4. Autor, Araceli Sebastián Ramos.

Fecha de emisión, 28 de abril de 1981. Hoy vamos a hablar de la validez de los test mentales. ¿Desde cuándo se utilizan los test mentales? La primera vez que aparece el término test mental en la literatura psicológica data de 1890 en un artículo de Cattell.

¿De qué trataba ese artículo? Cattell estudió allí el uso de los test de fuerza muscular, tiempo de reacción, rapidez de movimiento y otras medidas utilizadas entonces para determinar los niveles intelectuales de los estudiantes universitarios. ¿Hubo algún otro tipo de test que se aplicara por aquel entonces? Sí. Ebbinghaus y otros investigadores diseñaron también test de computación aritmética, memoria y terminación de oraciones que fueron utilizados para medir la inteligencia de los niños de primaria.

Naturalmente hubo otros muchos investigadores que aplicaron test importantes desembocando el esfuerzo común de todos en la escala de inteligencia de Biné. ¿Qué es la escala de Biné? Biné y Simón se asociaron en el año 1905 e idearon el primer test de inteligencia, semejante a los que actualmente se utilizan. Por supuesto que esta escala inicial carecía de un método objetivo para derivar una puntuación total, pero ya en 1908 apareció una segunda escala donde se introducía el término edad mental, que sirvió para comparar a niños retrasados con niños normales de la misma edad.

Tengo entendido que la Universidad de Stanford tuvo un papel importante en el comienzo de la elaboración de este tipo de test. ¿Es cierto? Sí. Hacia el año 1916 fue Thurman quien elaboró el llamado test de inteligencia de Stanford-Biné, en el que aparecía por primera vez el término de coeficiente intelectual.

¿Pero qué validez tiene un test? Un test es tan válido cuanto cumpla con fidelidad el fin para el que está destinado. ¿Pero si nosotros tenemos un test, cómo sabemos si tiene validez? En realidad la determinación de la validez de un test no suele ser una tarea muy sencilla. ¿Porque creéis que basta con la intención del autor del test de medir el aspecto o conducta que quiera evaluar? Yo creo que no, porque es posible que, a pesar de la capacidad y buena intención del autor, el test, en vez de medir lo que él se propuso, mida otra cosa diferente.

La cuestión radica pues en determinar hasta qué punto un test mide con suficiente precisión un determinado rasgo de conducta. Para lograrlo deberá existir una correspondencia entre las medidas de ese rasgo de conducta obtenidas a través de otros test y los resultados obtenidos con el test que queremos validar. Parece que la única manera de ver si un test es válido es comprobar que sus resultados son acordes con los de otros test con suficiente prestigio.

Pues sí, en realidad se trata de eso. Si bien, en última instancia, hay que elegir criterios de validez que no dependan del test en cuestión y que nos permitan hallar la correlación entre sus

resultados con el criterio que se haya seleccionado. Pero eso no debe de ser muy fácil, ¿no? A veces resulta sencillo seleccionar un criterio.

Por ejemplo, un buen criterio para valorar un test que pronostique el rendimiento de unos alumnos en una asignatura de un curso pueden ser las notas que en esa materia han obtenido esos mismos alumnos. Comparando los resultados del test aplicado con las notas obtenidas, que sería el criterio en este caso, podrán ocurrir dos cosas. O bien los resultados del test no están acorde con las notas en cuyo caso el test tiene poca validez, o al contrario.

Sí, pero ese es un ejemplo muy sencillo. ¿Qué ocurre cuando un test pretende valorar al mismo tiempo varios aspectos en conjunto, como ocurre cuando se hace una selección de personal, por ejemplo? Como es lógico, este caso es más complicado. La validez del test habría que determinarla en tal situación hallando la correlación con varios criterios independientemente o con alguna combinación de los criterios más relevantes del caso.

Naturalmente la dificultad radica ahora en la elección de un criterio o combinación lo más preciso y fiable posible, lo que es en esencia el problema más fundamental en la validación de un test en general. Estoy pensando que un test puede ser válido en cuanto a que el contenido de sus ítems estén relacionados con la finalidad del test, y que sin embargo puede que no resulte muy eficaz a la hora de predecir la conducta del sujeto en alguna situación específica. Por ejemplo, un test nos puede valorar muy bien los conocimientos de topografía de un guía de montaña respecto a su selección para ese puesto, y luego, en la práctica, a este guía puede no gustarle su trabajo por tener otra vocación.

¿Hasta qué punto podemos decir que ese test es válido si tratamos de seleccionar un buen guía? Ocurre que en realidad existen varios tipos de validez de un test. Siguiendo la clasificación de Anastasi diremos que hay validez de contenido, validez empírica y validez estructural. Sepamos en qué consisten.

La primera, validez de contenido, hace referencia a la relación del contenido con el objeto fundamental del test. Conociendo el tema objeto del test, lo que hay que analizar con detenimiento es el área de contenido para así poder seleccionar ítems que en la proporción adecuada comprendan los aspectos más fundamentales del tema. Pero esto en realidad se debe hacer siempre que se trate de elaborar un buen test, desde luego.

El análisis y la especificación clara de los objetivos que se pretenden alcanzar con el test es algo que debe preceder a la selección de aquellos ítems que se crean más acordes con los objetivos, si queremos que el test sea válido en cuanto contenido. ¿En qué tipos de test es más importante la validez de su contenido? Fundamentalmente en los test de rendimiento y para muchos tipos de test de medida, aunque con frecuencia, puede conducir a error. ¿Cómo es posible esto? En el caso de los test de aptitudes o de personalidad, puede ocurrir que el contenido del test tenga poco parecido intrínseco con los rasgos de conducta que se someten a prueba, por no basarse en un contenido específico o en experiencias uniformes previas que pudieran haber servido para la elaboración del test, ya que en este caso la forma de respuesta

de los individuos puede tener una gran variación.

¿Y la validez empírica en qué consiste? La validez empírica se calcula hallando la correlación entre la puntuación del test y un criterio directo e independiente de la finalidad del test. Volvemos otra vez al criterio. ¿Qué criterio hay que escoger en cada caso? Como ya dije antes, en la elección de un buen criterio radica la mayor dificultad y no siempre es fácil conseguirlo, hasta el punto que a menudo hay que recurrir a criterios con un valor relativo pero que resulten más fáciles de conseguir.

Por ejemplo, si se pretende validar test de inteligencia general, ¿qué criterios escogeríais vosotros que nos proporcionaran una buena información sobre el rendimiento académico en cada caso? Las calificaciones escolares. Quizás los premios obtenidos o las opiniones del profesorado. Pero estos no son criterios muy precisos, porque en cualquier caso sólo nos valorarían la inteligencia del sujeto en una época de su vida, y no nos dicen nada del desarrollo que posteriormente cada uno de los individuos haya podido dar a su intelecto.

Podría ocurrir que, según estos criterios, un alumno de primaria haya tenido una gran inteligencia en su infancia, pero que alumnos menos brillantes hayan llegado a superarles posteriormente por una actividad intelectual más continuada. Entonces, ¿en este caso serían necesarios criterios que controlaran mejor la actividad intelectual de los sujetos en épocas posteriores? En efecto. Sobre todo si lo que se pretende en el test como última meta es predecir cuál de estos sujetos es el que puede proporcionar mejor rendimiento.

¿Parecen estos últimos criterios más o menos precisos que los que propusisteis en principio? Son más precisos. Fijaos que este buen criterio es muy difícil de obtener, porque requiere una gran cantidad de tiempo de seguimiento del sujeto, y en cualquier caso es muy difícil controlar otros muchos factores, como pueden ser familia, sociedad, y otros que pueden condicionar posteriormente al sujeto. Creo que voy entendiendo la validez empírica.

¿Y la estructural en qué consiste? La validez estructural hace referencia a la validez de elaboración o construcción del test, contribuyendo a la comprensión del concepto elaborado que mide el test, de forma que nos permita vislumbrar claramente el campo de conducta que abarca el test. Naturalmente, cualquier información que contribuya a determinar y a esclarecer la naturaleza del rasgo de conducta que considera el test, y los condicionantes que repercutan en su desarrollo, contribuirán a determinar el ámbito de conducta que el test considera, y son por tanto, datos muy útiles para esta validez. Pero esta validez parece menos concreta que las otras.

Sí, pero en realidad, en un sentido amplio, comprende a los otros dos tipos de validez e incluye las técnicas señaladas para ello. Es pues más abstracta, pero puede ser más completa.