FF 17 22

Pedagogía experimental, título La validez de los test, autor, piedad, Granados García Tenorio, fecha de emisión 24 de febrero. La historia del test mental empieza en 1890 con un artículo de Cattel en el que trataba sobre los test de fuerza muscular, tiempo de reacción, rapidez de movimiento y otras medidas utilizadas entonces para determinar los niveles intelectuales universitarios. Naturalmente hubo otros investigadores, como por ejemplo Ewing Gauss, que diseñaron otros tipos de test, como los test de computación aritmética, de memoria y terminación de oraciones, que fueron utilizados para medir la inteligencia de los niños de primaria.

También hubo otros muchos investigadores que aplicaron test importantes desembocando el esfuerzo común de todos en la escala de inteligencia de Binet. ¿La escala de Binet tiene algo que ver esta escala con el coeficiente intelectual? Sí, Binet 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.

La universidad de Stanford tuvo un papel importante en el comienzo de la elaboración de este tipo de test. Hacia el año 1916 fue Thurman quien elaboró el llamado test de inteligencia de Stanford-Binet, en el que aparecía por primera vez el término de coeficiente intelectual. Consideramos válido un test en la medida en que éste cumple con fidelidad el fin para el que está destinado.

Pero si disponemos de un test, ¿cómo podemos saber si es válido o no? La determinación de la validez de un test no suele ser una tarea sencilla, ya que a pesar de la capacidad y buena intención del autor, puede ocurrir que en vez de medir lo que él se propuso, mida otra cosa diferente. Por tanto, la cuestión radica 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 la selección de los criterios no debe ser tarea muy fácil. 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 esas materias 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 acordes con las notas, en cuyo caso el test tiene muy 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 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 en esencia es el problema más fundamental en la validación de un test en general.

Un test puede ser válido en cuanto que el contenido de sus ítems estén relacionados con la finalidad del test, y, 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 poco o 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. Validez Estructural Sepamos en qué consisten los tipos de validez enunciados. La primera, es decir, la 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 a contenido.

Fundamentalmente en los test de rendimiento y para muchos tipos de test de medida, aunque con frecuencia puede conducir a error. En el caso de los test de aptitudes o de personalidad, puede ocurrir que el contenido del test tenga poco parecido e 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. En este caso, las formas de respuesta de los individuos pueden 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 de test. Como ya dijimos antes, en la elección de un buen criterio radica la mayor dificultad y no siempre es

fácil conseguirlo, hasta el punto de que a menudo hay que recurrir a criterios con un valor relativo, pero que resultan más fáciles de conseguir. Si pretenden validar test de inteligencia general, ¿qué criterios escogeríamos que nos proporcionaran una buena información sobre el rendimiento académico en cada caso? Tal vez 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 hubiera tenido una gran inteligencia en su infancia, pero que alumnos menos brillantes hubiesen llegado a superarle 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.

Estos últimos criterios son más precisos que los que propusimos en principio. Fijémonos 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 familiares, sociales y otros que pueden condicionar posteriormente al sujeto. Después de lo expuesto sobre la validez empírica pasamos a hacer algunas consideraciones sobre la validez estructural.

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. Naturalmente, cualquier información que contribuya a determinar y esclarecer la naturaleza del rasgo de conducta que considere 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. Esta validez parece menos concreta que las otras pero, en realidad, en un sentido amplio comprende a los otros dos tipos de validez a la validez de contenido y a la validez empírica e incluye las técnicas señaladas para ello es, pues, más abstracta pero puede ser más completa.