FF 17 07

ASIGNATURA PEDAGOGÍA EXPERIMENTAL TÍTULO FIABILIDAD DE LAS MEDIDAS EN PEDAGOGÍA GUIÓN Nº 3 AUTOR P. GRANADOS GARCÍA TENORIO FECHA DE EMISIÓN 10-03-81 Una pregunta que surge en relación con los exámenes es si se puede estar seguro de que un examen produce resultados consecuentes y, por tanto, es algo satisfactorio respecto a su fin. La realidad es que sí se puede estar seguro siempre que el examen sea fiable naturalmente. ¿Quiere decir que sea válido? No, no.

Hemos dicho fiable y no válido. La fiabilidad y la validez de un examen no son dos atributos recíprocos. ¿Cuándo podemos decir entonces que un examen es válido? Cuando de un modo demostrable mide lo que se pretendía medir con él.

Entonces, para que un examen sea válido, ha de ser fiable. Sí, pero puede ser fiable y no por ello ha de ser válido. Imagínate que estamos midiendo distancias con un metro defectuoso.

¿Tendría sentido registrar los resultados? No. ¿Por qué? Porque las medidas que hiciéramos carecerían de fiabilidad y, por tanto, sería imposible obtener con ellas una valoración válida. ¿Y si, en vez de tener un metro inútil, tuviéramos una báscula perfecta de gran precisión, tendríamos un instrumento fiable? Sí, claro.

¿Pero podríamos medir con este instrumento fiable las distancias que nos proponíamos? No. Pues bien, aquí tenemos un caso en que existe fiabilidad, pero no hay validez. ¿La fiabilidad y la validez hay que considerarlas entonces por separado? Sí.

Por supuesto, hay que asegurarse, en primer lugar, de que el examen produzca resultados consecuentes, pero no por ello podremos afirmar, en principio, que los resultados vayan a ser correctos, en el sentido de que el examen nos permita valorar aquellas características en las que estamos interesados. ¿Tan importante es tener en cuenta la fiabilidad a la hora de plantear un examen? Muy importante. Un profesor responsable debe tener en cuenta los factores que afectan a la fiabilidad, asegurarse de que las pruebas que emplea son las más satisfactorias posibles a este respecto y que pueda valorarse la fiabilidad de esas pruebas.

¿Realmente poner un buen examen es más difícil de lo que parece? ¿Cuáles son los factores que pueden afectar a la fiabilidad de un examen? Pues, por ejemplo, el contenido y la misma estructura del examen. ¿Cómo? Una pregunta imprecisa o ambigua que deje en duda a los alumnos respecto a la clase de respuesta que se espera de ellos reducirá la exactitud de nuestras evaluaciones. Pero entonces, por razones análogas, el examen en conjunto no puede ser fiable si no muestra de forma adecuada las capacidades que se trata de comprobar.

En efecto. ¿Y qué podemos deducir de ello? Que la fiabilidad aumenta en función del número de preguntas que contiene un examen. Es lógico.

Un examen que tuviera, por ejemplo, una sola pregunta, muy corta, deja en principio

demasiadas probabilidades a la suerte. E incluso podría darse el caso que esa pregunta fuera la única que no supiera un alumno o la única que supiera otro. Cuanto mayor número de preguntas contenga un examen, tanto más fiables llegarán a ser los resultados.

Sobre todo si esas preguntas son tan escogidas que nos permitan asegurar que se refieren a cada aspecto relevante de la materia de que se trata. ¿Y no dependerá la fiabilidad de los procedimientos de calificación que se utilice en cada caso? Sí. Hay muchas pruebas que demuestran que las calificaciones otorgadas basándose en una impresión subjetiva varían de un examinador a otro e incluso de una ocasión a otra para un mismo examinador.

¿Las preguntas de ensayo disminuyen entonces el grado de fiabilidad de un examen? Sí, a menos que se establezca expresamente un plan detallado y objetivo de calificaciones en tales pruebas. Entonces, ¿cuál es el sistema de calificación ideal? No puedo contestarte de un modo concreto. Pero, en principio, cualquiera que produzca resultados similares, independientemente de quién sea el profesor que califique los exámenes.

¿Existe algún otro factor que influya en la fiabilidad de un examen? Sí, hay otra serie de factores que escapan a nuestro control. En la medición educativa, tratamos de valorar el conocimiento y la capacidad que los alumnos han adquirido basándonos en su realización en una ocasión particular. Evidentemente, no tenemos garantías de que esta realización, aun cuando podamos valorarla con exactitud, haga plena justicia a las capacidades de sus respectivos intérpretes.

El grado en que un alumno demuestra efectivamente el conocimiento o la capacidad que posee depende de una gran variedad de factores personales, como son estado de ánimo, salud, motivación, etc. Nada o casi nada es lo que puede hacer un profesor para evitar que la fiabilidad de un examen resulte afectado por tales factores. Y lo que es peor, nunca podrá controlar los resultados de sus acciones en este sentido.

¿Pero esto hace poco fiables las mediciones de los exámenes? En algunos casos sí, aunque a veces es posible salvar esta dificultad o al menos amortiguar sus efectos repitiendo esas pruebas tras un corto intervalo y comprobando si los resultados son consecuentes. Las diferencias que surgen en estos casos no son imputables a defectos en el examen o errores de calificación, y muchas de ellas no serán sino el reflejo de cambios en las características que estamos valorando, porque si el intervalo de repetición de la prueba es breve, los alumnos no pueden haber tenido ocasión de incrementar notoriamente sus conocimientos, de modo que justifiquen esas diferencias. Preparar dos test o exámenes que sean estrictamente paralelos y que midan los mismos aspectos del conocimiento debe de ser difícil, ¿no? Sí, por ello se han buscado criterios de fiabilidad que pueden aplicarse a un solo test o examen.

En esto consiste, por ejemplo, el método de mitades, constituido por un test dividido en dos partes que están en correlación. Lo primero que hay que hacer en este tipo de métodos es asegurarse que las dos mitades sean comparables. Además, hemos de tener en cuenta el hecho de que la correlación que se obtiene de esta comparación se basa en test que son más cortos que el original y, como antes decíamos, la fiabilidad de un test depende de su longitud.

Cuanto más largo sea el test, mayor fiabilidad. Entonces, el método de mitades produce una estimación de fiabilidad por defecto. Sí, y esto a menos que empleemos una fórmula que compense el hecho de que estamos examinando un test que es exactamente la mitad de largo que el original.

La fórmula a aplicar sería fiabilidad de un test es igual a 2r dividido por 1 más r, donde r es la correlación medida entre las dos mitades del test. Por si no ha quedado claro, voy a repetir esta definición, esta fórmula. Fiabilidad de un test es igual a 2r dividido por 1 más r, 2r dividido por 1 más r, donde r es la correlación medida entre las dos mitades del test.

Profesor, estoy pensando que si la fiabilidad de una parte de un examen no corresponde a la fiabilidad de todo el conjunto, tal vez un examen pueda ser fiable, considerado todo entero, y no por ello podremos tratar como fiables a partes de ese examen. ¿Estoy en lo cierto? Sí. Podemos tener justificación para comparar las notas obtenidas por algunos alumnos en un examen o test como un todo.

Pero hay que ser más cautos a la hora de basar tales comparaciones en ciertas partes de ese test o examen, porque en este último caso estaríamos tratando con un nivel más bajo de fiabilidad. ¿Se puede valorar la fiabilidad de alguna otra manera? Existen otros métodos de valorar la fiabilidad de un examen, en general, más elaborados, como el método Kidder-Richardson, que basan la estimación sobre la clase de evidencia que brinda el análisis de preguntas. Pero estos ya son otros métodos que merecen un capítulo aparte.

Nuestra preocupación en cuanto a la fiabilidad es y debe ser hacer hincapié en la necesidad de asegurar que cualquier prueba o examen que se aplique sea todo lo fiable posible, aunque humildemente reconozcamos, e incluso establezcamos como base inicial, que la fiabilidad perfecta es inalcanzable. Subtítulos realizados por la comunidad de Amara.org