

Fiabilidad y generalizabilidad. Aplicaciones en la evaluación educativa

Carmen Díaz, Carmen Batanero y Belén Cobo

Introducción

Según Thorndike (1989), el proceso de medida en educación y psicología se propone ligar ciertos conceptos abstractos a indicadores empíricos. En el caso de la evaluación educativa, intentamos relacionar los conocimientos de los alumnos sobre un concepto, sus capacidades y estrategias en la resolución de problemas y tareas o sus actitudes con sus respuestas a los ítems de una prueba o de un cuestionario. El análisis de datos se hace sobre las respuestas, ya que son observables. El interés teórico, sin embargo, es el concepto subyacente (conocimientos, actitudes, etc.) que no podemos observar directamente, pero que tratamos de inferir a partir de las respuestas.

Cuando la relación entre los indicadores empíricos (respuestas) y los conceptos subyacentes es fuerte, el análisis de los indicadores nos permite hacer inferencias útiles sobre los conceptos teóricos y evaluar nuestras hipótesis previas sobre los mismos, así como tomar decisiones adecuadas sobre la acción didáctica. Para medir esta relación se han definido diversos conceptos, como validez, fiabilidad o generalizabilidad. Puesto que, cuando evaluamos a un alumno, podemos plantearle una infinidad de posibles preguntas sobre el mismo tema, e incluso con las mismas preguntas, las repuestas del alumno pueden variar dependiendo de su atención, cansancio u otros factores, debemos reconocer un carácter aleatorio a los resultados de la evaluación y, en general, de una investigación educativa.

Las fuentes de error en este proceso pueden ser de naturaleza determinista o aleatoria y afectan de diversa forma a nuestras conclusiones y decisiones. Los sesgos, de naturaleza determinista, aunque de magnitud desconocida, se derivan de nuestros procedimientos, tanto en la selección de la muestra, como en la elaboración de los instrumentos y en la toma o análisis de los datos. Por ejemplo, en un estudio de evaluación de conocimientos puede producirse un sesgo si la prueba de evaluación es demasiado sencilla (por lo cual los alumnos más brillantes no mostrarán toda su capacidad) o no contempla adecuadamente la variable que queremos medir (por ejemplo, si queremos evaluar la capacidad de resolución de problemas, en general, y sólo proponemos proble-