

Editorial:

Evaluación Estandarizada

Standardized Assessment

Jesús Miguel Jornet Meliá *

Universidad de Valencia

La estandarización de la medida y evaluación en CC. Sociales en general y en CC. de la Educación en particular ha estado en el centro de los debates epistemológicos tanto respecto a los procesos de construcción del conocimiento como en los aplicados, por ejemplo, la evaluación.

Por estandarización entendemos el proceso de sistematización de todos los elementos de acercamiento a una acción de recogida e interpretación de información, de manera que se utilicen los mismos: instrumentos o técnicas, criterios de corrección y/o síntesis o análisis de la información y criterios de interpretación de la misma.

Estas estrategias se pueden aplicar tanto para clasificar instrumentos y técnicas de recogida de información, como para clasificar planes o programas de evaluación. No obstante, su mayor impacto, sin duda, lo tiene en el diseño de instrumentos y/o técnicas de recogida de información.

Siguiendo esta posición, y respecto a instrumentos y técnicas, podemos identificar tres grandes tipos de estrategias:

- a) *Estandarización completa*. Implica que todos los elementos del instrumento de medida están sistematizados y se aplican de la misma manera para todas las personas. De este modo, se les presentan los mismos estímulos (reactivos o ítems), se les dan las mismas instrucciones de aplicación, se administran en un mismo tipo de situación (por ejemplo, si existe tiempo limitado o no), ellos tienen que ofrecer sus respuestas del mismo modo (por ejemplo, marcando o señalando en papel o seleccionando en ordenador, o escribiendo su respuesta), se corrigen de la misma manera y se interpretan las puntuaciones de acuerdo a los mismos criterios (estándares de interpretación, baremos, normas...).
- b) *Estandarización parcial*. Generalmente tienen prefijados o sistematizados los modos de recogida de información, si bien en la forma de sintetizar o analizar la información existen posibilidades alternativas o mayor flexibilidad.
- c) *Estandarización nula*. En este caso, ninguno de los elementos están prefijados y se recaba, sintetiza o analiza e interpreta la información de manera particular para cada caso.

La elección de uno u otro tipo de instrumento depende en muchas ocasiones de la posición epistemológica o metodológica del evaluador. No obstante, desde nuestro punto de vista, debería depender de las características del objeto de medición/evaluación y de

*Contacto: jornet@uv.es

issn: 1989-0397

www.rinace.net/riee/

<https://revistas.uam.es/riee>

la finalidad; de modo que nos situamos en una posición de complementariedad metodológica cuantitativa/cualitativa, reconociendo la complejidad de los hechos educativos y la necesidad de aunar esfuerzos desde diversas posturas para representar de manera adecuada la realidad objeto de estudio.

La estandarización ha estado defendida desde posiciones en las que se enfatiza la necesidad de plantear un mismo sistema de acercamiento a la realidad (sea en pruebas de rendimiento, escalas de actitudes o sistemas de evaluación de docentes, instituciones, programas...) con el fin de que las diferencias en relación a los resultados puedan imputarse a la diversidad del objeto o sujeto evaluado o a efectos debidos a una intervención sobre ellos, y no a posibles factores diferenciales derivados del modo en que han diseñado, aplicado o interpretado los instrumentos.

El supuesto de base es que debe mantenerse el conjunto de estímulos y/o modos de recogida de información de la misma manera con el fin de crear la misma situación. Las personas reaccionan ante la misma situación estimular de diversas maneras. Esa diversidad de respuesta caracteriza a las personas y no es debida a diferencias del instrumento o su uso. Este concepto, sin duda, es atractivo y parece lógico, pero la realidad nos demuestra que el hecho de estandarizar no siempre es adecuado y no asegura que se cumpla el supuesto mencionado, pues siempre existen factores que inciden de una u otra manera y que implican que cualquier medida contenga error, tal como señaló Spearman a comienzos del s. XX marcando el origen de la teoría clásica de construcción de tests.

Este planteamiento, sin embargo, ha sido característico de los procesos de recogida de información en los que, durante buena parte del siglo XX y el actual, se ha ido diferenciando entre instrumentos de medida, propiamente dichos, y las técnicas evaluativas. Los instrumentos de medida (tests psicológicos, pruebas de rendimiento, escalas de percepción y actitudes) están totalmente estandarizados, mientras que las técnicas evaluativas (cuestionarios, observación, entrevista...) pueden recorrer los tres niveles, si bien, lo más usual es que correspondan a procesos parcialmente estandarizados o con estandarización nula. En cualquier caso, podríamos imaginar un eje en el que en un extremo se situaran los instrumentos totalmente estandarizados -por ende supuestamente independientes del observador/evaluador-, y en el otro extremo, los no estandarizados -dependientes, en ese caso, de diversos factores, pero siendo central el rol del observador/evaluador-.

Todos los métodos y técnicas de recogida de información están sujetos a criterios de bondad: fiabilidad y validez. No obstante, por el tipo de metodología en que se sustenta cada procedimiento, la cuantificación de los criterios en indicadores que permitan informar de la fiabilidad de los mismos, o en evidencias que apoyen su validez, son más fácilmente identificables en los totalmente estandarizados. Ciertamente el control sobre los criterios de bondad de técnicas evaluativas resulta siempre más complejo, al no disponer de modelos de medida que los sustenten y por la necesidad de integrar informaciones más diversas, en las que el control sobre las mismas es más difícil de realizar. Ello no quiere decir que no posean tales características. Únicamente que resulta de mayor complejidad poder comprobarlas.

El problema de la validez, sin embargo, se identifica en todos los instrumentos y técnicas. Probablemente es el criterio más importante y a su vez más olvidado y que no se resuelve a través de modelos de medida. Recuérdese que la fiabilidad, en todo caso, es

una condición necesaria, pero nunca suficiente para la validez. Buena parte de los problemas que se imputan a las pruebas estandarizadas están relacionados precisamente con carencias en la validez. De ahí que objetivar sea una cuestión fundamental, que se enraíza en el sustrato de la validez y no queda nunca atendida por más que utilicemos los modelos de medición más avanzados. El problema es que, en muchas ocasiones, usuarios e incluso expertos creen que los modelos de medición aseguran todo y, por desgracia, no es así. En la base de muchos debates acerca de si la estandarización es una aproximación adecuada normalmente se identifican los problemas de fondo acerca de la validez.

Más allá de la evaluación de alumnado para procesos de selección o certificación, la estandarización también ha llegado a instaurarse en diversos procesos de evaluación, como por ejemplo la evaluación de docentes, instituciones y sistemas educativos. No obstante, si diseñar instrumentos de medida es complejo, establecer planes de evaluación (que deben integrar diversas informaciones, ser amplios y atender a diversas variables de entrada, proceso, contexto y resultado...) lo es más aún. Con todo, hay una tendencia clara en diversos tipos de planes –como por ejemplo, para evaluar instituciones o sistemas educativos– a reducir el concepto de calidad a la interpretación de los resultados del proceso educativo. En tales casos, es frecuente observar planes de evaluación que toman como eje vertebrador del análisis evaluativo el uso de pruebas estandarizadas de desempeño o logro educativos y, a lo sumo, integran otras variables recogidas a través de cuestionarios de contexto.

Sin pretender volver al debate sobre ello, pues nos hemos manifestado al respecto en diversos medios y ocasiones, es obvio que resultan planes muy parciales, en los que la validez suele estar en entredicho. Por ello, aunque la tendencia sea estandarizar también en este tipo de procesos, entendemos que más allá de la extracción de indicadores aislados de evaluación, la estandarización para situaciones más complejas (como son, por ejemplo, las instituciones o sistemas educativos, o la actuación docente...) es más que discutible. La dificultad de llevar el concepto de estandarización a programas o planes de evaluación radica, en definitiva, en las características del objeto de evaluación y en el concepto de calidad que quiera comprobarse. Por ello, a no ser que se reduzca a variables de resultado, es muy complicado intentar diseñar planes totalmente estandarizados. A lo sumo podrán integrarse algunos indicadores estandarizados en un plan que, por definición, deberá ser parcialmente estandarizado. Y, si se da lo anterior –reducir el plan a variables de resultado–, la validez de la evaluación será un factor deficitario, con toda seguridad.

En este número monográfico se integran diversos trabajos. Los tres primeros (de Agustín Tristán y Nancy Pedraza, Richard Phelps, y, finalmente, el de Manuel Fernández Navas, Noelia Alcaráz y Miguel Solá) presentan reflexiones de carácter teórico-metodológico que recogen diversas miradas de reflexión crítica acerca de la estandarización, desde posiciones más cercanas al análisis epistemológico hasta otras más impregnadas de ideología. En el siguiente artículo, Carolina Ruminot nos presenta un análisis sobre las consecuencias del SIMCE en Chile para el sistema educativo, basado en un análisis con docentes y proponiendo elementos derivados de ello para la intervención educativa. Pilar Rodríguez Morales nos muestra en su trabajo el diseño de una prueba de matemáticas y comprensión lectora para el ingreso de estudiantes de colectivos vulnerables en Uruguay, incluyendo aproximaciones de Teoría de Respuesta al Ítem. Una muestra de acercamiento de estandarización para evaluar la práctica docente, basada en observación y rúbricas y análisis sustentados sobre Teoría de la

Generalizabilidad, nos la aportan Luis H. Pedroza y Edna Luna, sin duda de gran interés para la valoración de las prácticas de aula. En los dos últimos trabajos, el de Cristina V. Hernández de la Toba y Joaquín Caso, y el de Carlos Sancho, Jesús M. Jornet y José González-Such, se presentan dos procesos de trabajo de estandarización sobre variables socio-afectivas. En el primer caso, sobre una escala de convivencia escolar y, en el segundo, sobre otra orientada a analizar el valor social que subjetivamente se da a la educación. En conjunto, se ofrece un recorrido sobre trabajos en torno a la estandarización que esperamos sean de interés. Agradecemos, desde estas líneas, a los autores por su trabajo, así como a la Revista Iberoamericana de Evaluación Educativa por abrir este tema a debate.