

# El concepto de validez de los procesos de evaluación de la docencia

## *The concept of validity of teaching evaluation processes*

Dr. Jesús Miguel JORNET MELIÁ. Catedrático. Universitat de València ([jornet@uv.es](mailto:jornet@uv.es)).

Dra. María Jesús PERALES MONTOLÍO. Profesora Titular. Universitat de València ([perales@uv.es](mailto:perales@uv.es)).

Dr. José GONZÁLEZ-SUCH. Catedrático. Universitat de València ([jose.gonzalez@uv.es](mailto:jose.gonzalez@uv.es)).

### Resumen:

En este trabajo se presenta una propuesta conceptual y metodológica sobre la validez en los procesos de evaluación de la docencia. Pese a ser procesos muy habituales, y en algunos casos con significativas consecuencias individuales y organizacionales, no se ha realizado todavía una aproximación teórica y metodológica sobre la validez de los mismos, siendo este el objetivo de este trabajo. A partir de la revisión bibliográfica y de la experiencia acumulada durante años realizando evaluación e investigación evaluativa, tomamos como punto de partida una conceptualización propia de la evaluación (sumativa y formativa o valoración/diagnóstico) como una modalidad de investigación aplicada, que utiliza de forma esencial procesos de medición. Por ello, toma el concepto de validez en estos dos ámbitos de conocimiento y los lleva al área de la evaluación de la docencia, aportando una definición específica

de validez de la evaluación de la docencia. Además, a partir de los referentes conceptuales en investigación y en medición, concreta las facetas metodológicas, genéricas y específicas que integran ese concepto de validez, aportando definiciones propias para cada una de ellas. Finalmente, entendiendo que la validez, como calidad de la evaluación, debe ser cuidada durante todo el proceso evaluativo, partimos de su definición y de las facetas que la constituyen y proponemos una serie de estrategias para poder asegurarla de forma adecuada. Por lo tanto, la aportación de este trabajo es de carácter conceptual y metodológico, y propone un punto de referencia para conceptualizar la validez en un ámbito de amplio uso y con profundas consecuencias para los profesores, las instituciones y los sistemas educativos.

**Descriptor:** evaluación, profesorado, validez, investigación, medición.

---

Fecha de recepción de la versión definitiva de este artículo: 21-12-2019.

Cómo citar este artículo: Jornet Meliá, J. M., Perales Montolío, M. J. y González-Such, J. (2020). El concepto de validez de los procesos de evaluación de la docencia | *The concept of validity of teaching evaluation processes*. *Revista Española de Pedagogía*, 78 (276), 233-252. doi: <https://doi.org/10.22550/REP78-2-2020-01>

<https://revistadepedagogia.org/>

ISSN: 0034-9461 (Impreso), 2174-0909 (Online)

## Abstract:

This paper presents a conceptual and methodological proposal about validity in teaching assessment processes. Although they are very common processes, and in some cases have significant individual and organisational consequences, a theoretical and methodological approach to them has not yet been performed. Based on reviewing research and the experience accumulated over years of conducting evaluations and evaluative research, we take as a starting point our own conceptualisation of evaluation (summative and formative evaluation or appraisal/diagnosis) as an applied research form that essentially uses measurement processes. Therefore, it takes the concept of validity in these two knowledge areas and applies it to the area of teaching evaluation by providing a specific definition of the validity in the eva-

luation process. Moreover, based on conceptual references in research and measurement, this work identifies generic and specific methodological considerations, which include the concept of validity, providing meaningful definitions for each of them. Understanding that validity, as a quality of evaluation, must be taken into account throughout the evaluation process, we also propose a series of strategies to ensure this is done correctly, based on defining validity and the aspects that comprise it. Therefore, this work's contribution is conceptual and methodological, and it proposes a reference point for conceptualising validity in a field with broad use and with profound consequences for teachers, institutions, and educational systems.

**Keywords:** evaluation, faculty, validity, research, measurement.

## 1. Introducción

La evaluación de la docencia se ha ido implantando progresivamente como apoyo a la gestión, innovación y mejora de la calidad educativa. En educación superior es donde mayor número de experiencias se encuentran con consecuencias de rendición de cuentas (García Mestanza, 2010; Tejedor y García-Valcárcel, 2010). En los niveles preuniversitarios estos usos están teniendo mayor impacto en los últimos años (Guerra García, Leyva Barajas y Conzuelo Serrato, 2019; Manzi, González y Sun, 2011; Nava y Rueda, 2014). Sin embargo, cuando se revisa la literatura referida a la metodología de evaluación de la docencia o del profesorado,

se aprecian carencias importantes en el estudio acerca de la validez de las evaluaciones.

En este trabajo nos aproximaremos a algunas reflexiones de tipo metodológico que coadyuven a revisar los aspectos implicados en la validez de la evaluación de la docencia. Antes de adentrarnos en este análisis, estimamos necesario realizar algunas puntualizaciones. En primer lugar, respecto al uso de los términos *evaluación de la docencia* o *del profesorado*, entendemos que en ocasiones se diferencian los enfoques. El término *docencia* se utiliza preferentemente en aquellas orientaciones que asumen que la actuación del profesorado se da en un

contexto institucional, en definitiva, que lo individual no puede analizarse de forma independiente del contexto en que se trabaja. Por su parte, cuando se utiliza el término *evaluación del profesorado* el énfasis se sitúa en la valoración de la actuación profesional. Nosotros, precisamente por tratarse de un artículo sobre validez, los asumimos como sinónimos. No podemos entender la valoración profesional del profesorado descontextualizada del ámbito institucional en el que actúa.

En segundo lugar, tenemos que señalar que esta posición no niega que las diferentes tipologías de profesorado (por niveles educativos, materias, etc.) y contextos en que se desarrollan (modalidades de enseñanza —presencial/a distancia—, tipos de instituciones —públicas/privadas—, contextos en que se desarrolla la docencia —vulnerables vs. sobre-estimulados—, etc.) requerirían un análisis particular acerca de los factores involucrados en la validez de su evaluación, pues en cada caso particular, si deseamos evaluaciones de calidad, es necesario que los procesos de evaluación reconozcan estas particularidades. Sin embargo, en este trabajo, nos vamos a mover en un planteamiento genérico, aún a riesgo de dejar de lado concreciones necesarias, las cuales esperamos realizar en análisis posteriores referidos a tipologías de docencia particulares.

Esta revisión la estructuramos en los siguientes apartados: el concepto de validez en los procesos de evaluación, las facetas del concepto de validez y las estrategias de consecución de la validez.

## 2. Del concepto de validez

El concepto de validez ha sido centro de atención como criterio de bondad tanto de los procesos de investigación básica en ciencias sociales y de la educación, como del diseño de instrumentos de medición y evaluación. Así, la validez, en cualquiera de los ámbitos que afectan a la medición y la evaluación en las ciencias de la educación, constituye un tema central no solo para la construcción del conocimiento científico, sino también y especialmente en relación con la adecuación de la toma de decisiones y el ajuste de las propuestas de intervención.

El concepto de validez es multidimensional y heteróclito pues puede adoptar diferentes formas al aplicarlo a diversas unidades u objetos de investigación, medición o evaluación educativas. Los requisitos de validación no son uniformes y dependen, en gran medida, del ajuste global del proceso a las necesidades, fines y características del objeto. Por ello, un planteamiento del tema en el contexto de la evaluación en educación requiere de una cierta graduación de los conceptos, teniendo en cuenta que, si bien podemos identificar elementos comunes de aplicación a cualquier situación, también la atención a las características específicas de cada una de ellas puede conllevar énfasis diferenciados en cuanto a las acepciones o enfoques particulares implicadas en la validez.

Partimos de un concepto de validez en la evaluación que venimos desarrollando desde hace varios años y que formulamos inicialmente para la evaluación de programas como disciplina (González-Such, 1988; Jornet Meliá, Suárez Rodríguez y Pérez Carbonell, 2000; Ruiz-Primo, Jornet Meliá

y Backhoff, 2006; Jornet Meliá et al., 2017) y para la evaluación de la docencia (Jornet Meliá, 2014). Asimismo, esta concepción proviene de la experiencia acumulada a partir de diversos proyectos de investigación (UNIVECS [EDU2016-78065-R], SECS/EVALNEC [EDU2012-34734], EVALEF [EDU2011-29467], MAVACO [EDU2009-13485] y AVACO [SEJ 2005-05 923]), desarrollados por nuestro grupo de investigación GemEduco (<https://www.uv.es/gem/gemeduco/>) y otros (Rueda-Beltrán, 2019), en colaboración con la Red Iberoamericana de Investigadores en Evaluación de la Docencia – RIIED (<http://iide.ens.uabc.mx/ried/ibero/>).

Aunque se ha intentado trasladar el concepto de validez desde la investigación y la medición a la evaluación, consideramos que no hemos sido capaces todavía de hacer una adaptación que satisfaga las necesidades de conceptualización de la calidad —o validez— de la evaluación.

Desde posiciones post-positivistas de evaluación, se aportan los conceptos propios de la validez que se dan en la medición y en la investigación básica. Si asumimos que la evaluación es una modalidad de investigación aplicada, aún con dificultades conceptuales, la traslación del modo en que se entiende la validez se podría realizar parcialmente (Delgado Álvarez, 2014). Somos conscientes de que esta posición, asumir la evaluación como una modalidad de la investigación aplicada, no es compartida por todos los autores. No obstante, nosotros entendemos la evaluación como una modalidad de la investigación aplicada en educación en tanto en cuanto:

- a) Aporta información sobre el fenómeno educativo en contextos naturales, sin manipulación de variables.
- b) Permite identificar y diferenciar características vinculadas a prácticas adecuadas y deficientes, así como a factores que pueden dinamizarse para la mejora de la educación en un contexto determinado.
- c) La evaluación, si se asume como proceso de comprensión de la realidad educativa, cuanto menos guarda características de una «mirada inductiva del hecho educativo», por lo que permite identificar elementos que coadyuven a la generación de preguntas de investigación, el planteamiento de objetivos e hipótesis y, en consecuencia, la proposición de posibles teorías, emanadas del análisis directo de la realidad educativa.

Por otra parte, desde una posición de evaluación cualitativa, el concepto de validez se concreta en estrategias de acercamiento tales como la multiplicidad de fuentes e instrumentos y la triangulación de informaciones (Sandín Esteban, 2000).

En ambas aproximaciones, post-positivistas y cualitativas, las lógicas subyacentes son coincidentes en bastantes casos. Por ejemplo, nótese que la triangulación es una estrategia de razonamiento lógico similar a la de la validación criterial de instrumentos. En el caso de la triangulación, la diferencia respecto a la validez criterial de instrumentos radica en que no se asume que exista un criterio de verdad contra el que se juzga la calidad de la información recabada por el proceso evaluativo a validar, sino que se contrastan

tres o más fuentes o instrumentos buscando la existencia de consenso intersubjetivo (Jornet Meliá et al., 2017; Jornet Meliá, Bakieva y Sánchez-Delgado, en prensa).

De este modo, considerando que las referencias son escasas y están normalmente alejadas de casos particulares de evaluación de la docencia, no resulta fácil hacer una buena adaptación del concepto de validez para este ámbito. Por ello, orientamos nuestra reflexión en la adaptación del concepto de validación de la evaluación de la docencia partiendo de dos referentes fundamentales: los conceptos usuales de validez en la investigación básica y las diferentes acepciones de validez utilizadas en medición.

Revisamos a continuación los elementos que condicionan el concepto de validez:

1. *El concepto de evaluación.* Es un proceso sistemático de recogida de información orientado a la indagación para *comprender* la realidad de forma que podamos apoyarnos en estas informaciones para tomar decisiones que permitan la mejora de la intervención educativa.
2. *La función docente es una acción sistémica.* En el caso de la evaluación de la docencia, excede la comprensión de la intervención específica que el profesorado pueda realizar en el proceso de enseñanza-aprendizaje, asumiendo que la actuación docente es una acción global. Depende del contexto en que el profesorado trabaje, de forma que sus funciones pueden ser muy variadas y, en todo caso, deberán res-

ponder a las características de la cultura y realidad institucional. Por ello, un componente básico de la validez de la evaluación de la docencia será que el diseño del modelo de evaluación esté adaptado a la realidad institucional. *La evaluación debe realizarse en ambientes naturales.* A diferencia de la investigación básica, no podemos manipular la realidad para asegurar la calidad de la interpretación de resultados (Tejedor, 2004; Sánchez-Gómez, Martín-Cilleros, Costa y Peñalvo, 2018). Esto es así porque se evalúa en contextos reales, intentando que la evaluación no introduzca factores que modifiquen la realidad. En este caso, la validación de la evaluación se asemeja a la validación de la observación natural (Anguera, Blanco, Losada y Portell, 2018). En consecuencia, una condición fundamental para la validez de la evaluación será que se haya recogido la información necesaria en una situación real, incluyendo materiales de trabajo realizados o utilizados en la docencia (Lorenzo, 2011; Ortega Gaite, Tejedor Mardomingo y Perales Montolío, 2019), sin modificar la situación que se desea evaluar.

3. *La realidad es dinámica y cambiante.* Por ello, cuando evaluamos procesos educativos en general y docentes en particular, hay que tener en cuenta que la evaluación debe incluir en su planificación elementos que flexibilicen y faciliten su adaptación para poder aprehender los cambios que se den a lo largo de todo el proceso. En consecuencia, la validez deberá ser

un aspecto que contemplar a lo largo de todo el proceso de evaluación. Por ello, se podría analizar cuáles son los factores implicados en la validación del diseño, de la implementación, de los resultados, de su interpretación y del uso de la información. No obstante, en este trabajo, nos referiremos al proceso global.

4. *La orientación epistemológica y el rol del evaluador/a deben someterse a las características de la situación que se va a evaluar.* Por ello, nuestra posición evaluativa, que condicionará también el concepto de validez, es la de complementariedad metodológica o modelos mixtos cuantitativos/cualitativos (Bericat Alastuey, 1998; Mertens, 2018; Sánchez-Gómez, Rodrigues y Costa, 2018)
5. *La evaluación debe ser útil.* La evaluación tiene sentido como guía para la gestión educativa o elemento de mejora de cualquier proceso implicado (enseñanza-aprendizaje, docente, orientador, etc.). Solo podemos considerar que la evaluación es válida si la información que se extrae a partir de ella es útil para dinamizar la mejora y/o innovación de la educación. En este sentido, las particularidades de cada tipo de evaluación (sea por finalidad —formativa/sumativa— y/o por objeto de evaluación —tipología de profesorado, contexto, etc.—) deben ser respetadas en los procesos de evaluación (Martín-García, Sánchez-Gómez y Costa, 2019; González-Sanmamed,

Muñoz-Carril y Santos-Caamaño, 2019; González-Sanmamed, Sangrá, Souto-Seijo y Estévez, 2018). La información que se recoja para todo el proceso evaluativo debe estar conectada con el tipo de decisiones que se pretenden apoyar en ella.

6. *La evaluación debe realizarse con los mismos requisitos de rigor que son exigibles en la medición o en la investigación básica.* Evaluar requiere trabajar sobre la realidad sin modificarla. La ventaja, frente a la investigación básica, es que no se pretende obtener resultados cuyas interpretaciones sean generalizables para construir el conocimiento científico (Tejedor, 2004), mientras que el inconveniente frente a ella es que no podemos manipular la realidad para asegurar que la interpretación de resultados sea indudable.

En síntesis, considerando los elementos anteriormente señalados, podríamos acercarnos al concepto de validez en la evaluación de la docencia como el grado de representatividad y utilidad que nos permite la interpretación de los resultados de la evaluación, considerando que se hayan atendido todos los elementos del perfil docente de referencia, en la situación o contexto en que se da la docencia, sin interferir en la realidad del proceso educativo que estamos evaluando.

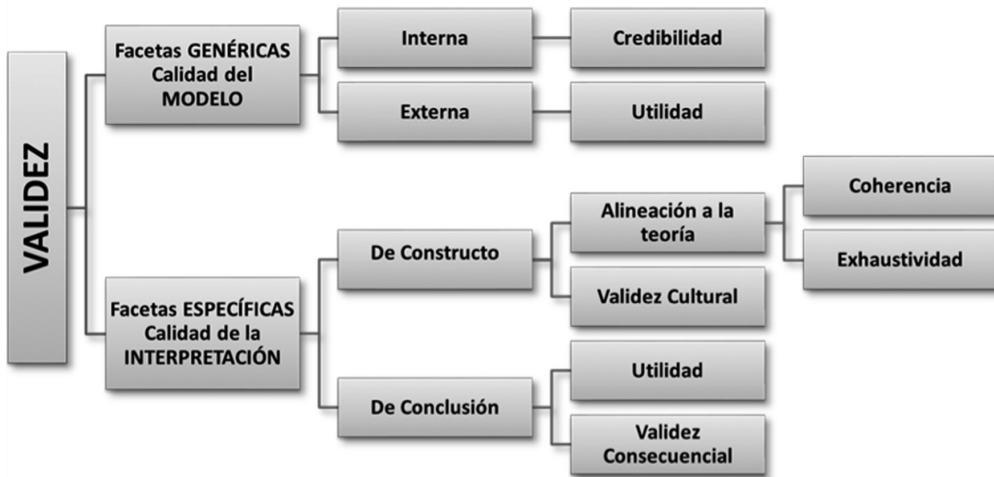
En definitiva, se trata de asegurar la calidad del proceso de evaluación educativa, comprobando si los componentes metodológicos implicados permiten:

- Representar de forma realista las características de la unidad objeto a la que se dirige.
  - Facilitar la consecución de los fines establecidos para el proceso de evaluación educativa.
  - Respetar los intereses de todos los implicados o afectados por el proceso evaluativo.
  - Representar el contexto en el desarrollo del proceso evaluativo, en la modulación de la interpretación de resultados y en la adecuación de sus consecuencias.
- Y coadyuvar a la mejora o innovación de la unidad objeto de evaluación.

### 3. Facetas del concepto de validez

Utilizamos el término *facetas* en relación con las diferentes acepciones y/o enfoques que pueden identificarse en el concepto de validez y las llamamos metodológicas porque se derivan de los componentes metodológicos clásicos de calidad de la investigación y la medición educativas. Entre ellas, se pueden diferenciar los tópicos (Rutman, 1984) que provienen de la investigación (facetas genéricas) y los que tienen su origen en la medición (facetas específicas) —ver Gráfico 1—.

GRÁFICO 1. Clasificación de las facetas del concepto de validez en la evaluación.



Fuente: Elaboración propia.

Las *facetas genéricas de la validez* provienen de posiciones post-positivistas en la investigación y se vinculan con la calidad del *modelo* de evaluación. Al aplicarlas para definir acepciones de validez en la evaluación, podemos señalar que están

presentes en cualquier proceso, si bien su importancia depende de las características de la unidad objeto de evaluación y de la situación evaluativa. Entre ellas, diferenciamos entre validez interna y externa.

Por *validez interna* nos referimos al conjunto de criterios de calidad que se orientan por los principios de *consistencia* de los diferentes elementos que componen el proceso la evaluación. Un buen nivel de validez interna de un proceso de evaluación de la docencia incluye como requisitos:

- a) Existencia de *coherencia interna* entre los indicadores base para informar de todo el proceso evaluativo y la posterior toma de decisiones.
- b) *Exhaustividad* en la selección de criterios e indicadores que se utilizan en la evaluación, es decir, cuando la evaluación cumpla la característica de generalidad de la representación del objeto.
- c) *Adecuación* del tipo de información incluida en la evaluación para satisfacer el *propósito y finalidad* de la misma.

Consideramos que la validez interna, es la base de la *credibilidad* de la evaluación. Es difícil que una evaluación sea creíble si no cumple con las características anteriormente descritas. Por ejemplo, en muchos sistemas de evaluación de la docencia se pretende evaluar al profesorado tomando como referencia únicamente los resultados de sus estudiantes. Es obvio que este tipo de evaluaciones incumplen los tres criterios anteriormente descritos y, por ello, ante la falta de credibilidad, son evaluaciones fuertemente rechazadas. Por el contrario, si el proceso de evaluación de la docencia se basa en una estructura coherente y exhaustiva

de indicadores, que representan el perfil del profesorado que se considera adecuado al contexto en que se desarrolla la docencia, de forma que permita responder al por qué y al para qué de la evaluación planteada, hay muchas más opciones de que esta evaluación sea creíble para los colectivos implicados y por tanto tenga validez interna.

Por otra parte, consideramos que una evaluación tiene *validez externa* cuando la información recabada y el modo en que se ha guiado todo el proceso evaluativo sirve para el uso que se había previsto para la evaluación. Es decir, se trata de asegurar que el conjunto del proceso evaluativo, desde el diseño del mismo, pasando por la información recogida y el modo de analizarla y llegando hasta el modo de transmitirla a los colectivos implicados, se adecua al principio de *utilidad* que animó la puesta en marcha de la evaluación. En este sentido, entendemos que la validez externa es en gran medida consecuencia de la validez interna, es decir, no se puede conseguir un buen nivel de validez externa en la evaluación si no se ha cuidado especialmente su validez interna. De este modo, no podemos pretender que la evaluación sea útil si no se ha garantizado previamente su credibilidad. No obstante, asegurar la validez interna es una condición necesaria, pero no es suficiente para asegurar la validez externa<sup>1</sup>. Así, por ejemplo, puede darse la situación de que un proceso de evaluación del profesorado tenga un gran nivel de validez interna para una evaluación formativa, pero si se comienza a utilizar —tal cual fue diseñado— para usos sumativos pierde totalmente la validez externa. Ese tipo de casos es frecuente.



En síntesis, validez interna y externa podrían asumirse como referencias para la credibilidad y la utilidad de la evaluación de la docencia y, en conjunto, forman un círculo virtuoso en el que cada una de ellas retroalimenta la adecuación de la otra: si una evaluación no es creíble no puede ser útil, pero es tanto más creíble cuanto más útil es. La validez interna, entendida como solidez del modelo de evaluación y base de la credibilidad de la evaluación, depende en gran medida del evaluador y del proceso de evaluación que ha diseñado; pero en la validez externa, entendida para utilidad de la evaluación y vinculada con los usos que se hacen de ella, pueden influir muchos factores espurios que introduzcan sesgos que no hayan sido previstos por el evaluador (Vedung, 1995; Weiss, 1998).

Al definir las facetas metodológicas de la validez distinguimos, como hemos indicado, entre facetas genéricas y facetas específicas. Nos referimos a las *facetas específicas* como una reinterpretación para la evaluación de enfoques de validez derivados de la medición, y que referimos en el caso de la evaluación de la docencia a la calidad de la *interpretación* de resultados —ver Gráfico 1—. Incluimos la validez de constructo y la validez de conclusión.

La *validez de constructo* hace referencia al grado en que todos los indicadores incluidos en el plan de evaluación responden al planteamiento teórico que orienta la definición de calidad de lo que queremos evaluar, de forma que permitan una interpretación de puntuaciones alineada a la teoría. Como en el caso de los instrumentos estandarizados de medición, se integra lo que denominaría-

mos validez de contenido, en tanto en cuanto los indicadores seleccionados deben permitir una visión *coherente y exhaustiva* del conjunto del fenómeno a evaluar. Por ejemplo, supongamos que se desea evaluar la docencia que lleva a cabo el profesorado en un centro de orientación pedagógica Montessori. Obviamente el perfil de la actuación docente estará delimitado por el modo en que desde el enfoque Montessori se deba actuar e incluirá todos los elementos que formen parte de la acción docente para que sus funciones dentro de la institución sean adecuadas. No valdría, por tanto, por interesante que pueda resultar, utilizar un enfoque evaluativo derivado del Modelo Europeo de Gestión de Calidad —EFQM— o cualquier modelo que la administración educativa arbitre para ser utilizado en su demarcación geopolítica<sup>2</sup>.

Asimismo, incluimos la *validez cultural* como una dimensión de la validez de constructo. Considerando que la docencia, y cualquier hecho educativo, se da en un contexto socioeconómico y cultural determinado, la selección de indicadores para la evaluación, así como su operativización en instrumentos de medida o en técnicas evaluativas de recogida de información, deben estar convenientemente adaptados al contexto. En ciencias sociales y educativas, la teorización nace de un contexto y por ello no siempre es generalizable. Por ejemplo, un modelo de evaluación de centros escolares de educación primaria diseñado según el Modelo Europeo de Gestión de Calidad (EFQM) puede incluir un perfil docente que integre tareas de enseñanza-aprendizaje, innovación y gestión. Entre estas últimas, es posible que se integren competencias relativas al trabajo docente colaborativo o en equipo o al liderazgo

académico. Sería imposible aplicar este tipo de modelo en centros escolares multigrado rurales en los que actúa un único docente. Allí, el perfil docente debería incluir otras competencias entre las que predominen las de autonomía personal y atención a la multiplicidad de hechos que en ese caso dependerán de un solo docente. Otros ámbitos son los relativos a la adaptación lingüística y cultural de instrumentos (Solano-Flores, Contreras-Niño y Backhoff, 2006; Solano-Flores y Milbourn, 2016), que, en ocasiones, no se atiende y provoca un uso indebido de los mismos. Debe tenerse en cuenta que las desadaptaciones son fuentes de sesgo y, en consecuencia, afectan directamente a la interpretación de resultados con pérdida de la validez.

La segunda faceta específica es la *validez de conclusión*, referida a si las *conclusiones* a las que se puede llegar a través de la evaluación *sirven al propósito* y finalidad que justificaron su puesta en marcha. Es decir, se trata de asegurar que la interpretación de resultados permite dar respuesta a los motivos y finalidades que justificaron el uso de la evaluación, sea como instrumento para la mejora o para la rendición de cuentas educativas.

Desde el ámbito de la investigación se incluyen como características de la validez de la conclusión: la consistencia, la generalidad, la relevancia y la utilidad. Entendemos que están intrínsecamente implicadas. No obstante, es necesario resaltar que pueden sintetizarse en dos grandes acepciones que se han desarrollado especialmente en el ámbito de la evaluación: la utilidad y la validez consecuencial. La *utilidad*, en este caso, es la referencia pragmática de la validez externa,

es decir, respondería a la pregunta ¿la interpretación de resultados de la evaluación satisface la finalidad de la evaluación? Por este motivo, deberá ser ajustada a la tipología de docencia, contexto o situación y a la finalidad de la evaluación (formativa/sumativa) concretas a la que se refiera la evaluación sobre la que analicemos su validez.

Finalmente, y de forma complementaria a la utilidad en la validez de la conclusión, la *validez consecuencial* es una acepción que precisamente ha irrumpido con fuerza en el ámbito de la evaluación (Martínez Rizo, 2015; Taut y Palacios, 2016). Básicamente, hace referencia a si el hecho de haber realizado una evaluación ha conllevado las consecuencias deseadas o no. El desajuste en cuanto al logro de consecuencias de cualquier evaluación puede darse por defecto o por exceso. Por ejemplo, la evaluación censal de los centros escolares de educación obligatoria de una región se ha podido llevar a cabo para mejorar la docencia en cada uno de ellos. Incluso, se podría haber explicitado que trata de desarrollar la cultura de evaluación para mejorar, por lo que se diseña un tipo de evaluación orientada a la identificación de buenas prácticas educativas, lo que redundaría en la mejora de la docencia a nivel regional. En el plan de evaluación, más allá de considerar algunas variables relacionadas con el nivel socioeconómico y cultural de las familias que atiende cada centro escolar, se podrían haber considerado otras relacionadas con la titularidad de los centros, su tamaño, etc. Y, como indicadores clave sobre los que se apoya la evaluación de la calidad, podría haberse tomado como referencia el nivel de logro del alumnado de cada centro mostrado en

pruebas estandarizadas en diversas materias. La información recogida se aportaría a cada centro escolar junto con la media global de los centros equivalentes, para que desde el propio centro se pueda apoyar una autoevaluación. Se asegura, además, que no se difundirán los resultados. Así planteado, el plan de evaluación obviamente es deficiente, pues no incluye suficientes variables de entrada y contexto y ninguna de proceso, presentando, por tanto, deficiencias en su validez interna y, como consecuencia de ello, carencias en cuanto a su validez externa y utilidad. Sin embargo, se puede decir que tiene suficiente validez consecucional porque, en parte, puede satisfacer el propósito y finalidad de la evaluación. Demos un paso más en el ejemplo: en una provincia concreta, al concluir el proceso, algún implicado con acceso a toda la información «filtra a los medios de comunicación» un ranking de centros ordenados por el promedio alcanzado por sus alumnos en cada materia. El uso de ordenaciones (ranking) no estaba previsto, ni tampoco que se difundiera públicamente la información. El hecho de que esta información fuera tratada así y se difundiera de ese modo atenta claramente contra la validez consecucional de la evaluación, deja de ser útil y pierde la validez de la conclusión.

Estimamos que estas facetas, tanto las genéricas como las específicas, pueden constituir un marco de referencia conceptual que permite orientar el análisis de validación de cualquier plan de evaluación de la docencia, desde su diseño previo hasta la valoración de su validez final.

## 4. Estrategias de consecución de la validez

Además de hacer la reflexión teórica sobre el concepto de validez en evaluación, es necesario aportar los elementos prácticos que permitan cuidar que efectivamente la evaluación sea válida. Se organizan estos elementos prácticos en torno a dos tipos de estrategias: las vinculadas con la planificación de la evaluación, y las relativas a la gestión de la información.

### 4.1. Planificación de la evaluación

La planificación de la evaluación es fundamental como pilar de su validez (Jornet Meliá, Suárez Rodríguez y Perales Montolío, 2003) y para llevarla a cabo de una manera adecuada debe contemplar todos los componentes del diseño y gestión de la evaluación. Estos componentes son:

- Definición del objeto y la finalidad de la evaluación. Se trata de definir todos los elementos que pueden afectar al objeto evaluado, para garantizar que la evaluación es una respuesta adecuada para aquello que se pretende. Para evaluar al profesorado es necesario tener muy claro el referente de perfil de calidad que se va a tomar en cuenta, y este no es generalizable, sino muy dependiente del nivel en que se trabaja (infantil, primaria, secundaria, postobligatoria, universidad) y de la modalidad de trabajo (presencial o a distancia/virtual), dado que el perfil de calidad es muy dependiente de las funciones docentes. La definición de la unidad objeto es el elemento clave inicial de validación, porque trata de definir las funciones, competencias,

habilidades y actitudes que debe tener el profesorado para ser considerado de calidad. En este sentido, por ejemplo, establecer el peso diferencial de la competencia emocional en la evaluación del profesorado de infantil, o de las competencias de tutorización en el profesorado universitario, y cuál es la fuente de información más pertinente en cada caso (familias o el propio estudiantado) es fundamental en la definición de la unidad objeto. Por tanto, una definición completa, coherente y exhaustiva del modelo de evaluación, que recoja todas las dimensiones implicadas reconocidas por el marco teórico con respeto al contexto cultural donde se realiza, contribuye a garantizar la validez interna de la evaluación, y por tanto, su credibilidad.

- En segundo lugar, es necesario anticipar facilitadores y obstáculos en el proceso de evaluación. Es muy importante cuidar el proceso de mediación<sup>3</sup> en la evaluación, con la finalidad prioritaria de promover la aceptación de la evaluación, siendo claro también en el establecimiento de los límites de uso de la información evaluativa, y ajustarlo a la finalidad de la evaluación de la docencia de que se trate. En todo caso, la evaluación siempre va a ser un acto de mediación entre todos los colectivos implicados (docentes, alumnado, familias autoridades y/o gestores-administradores de la educación, sociedad...), según el nivel educativo de que se trate y el uso final que se quiera hacer de los resultados

de la evaluación. Esto es especialmente importante en las evaluaciones de la docencia con consecuencias sumativas (ingreso, mantenimiento, promoción). En este sentido, los componentes fundamentales que debemos plantear en esta especificación de la evaluación es determinar qué se va a evaluar (en términos concretos) y los usos de la evaluación. Por ello, llegamos a presentar, en los años ochenta del s. xx, la propuesta de definir un *contrato de evaluación*, que permitiera establecer de forma explícita y consensuada la finalidad de la evaluación, la participación que van a tener los diferentes colectivos implicados, el tipo de información recogida, los análisis que se realizarán de la misma, y el uso que se hará de los resultados, concretando la forma de difusión, la confidencialidad o no de cada tipo de datos y los receptores de los informes.

- En tercer lugar, y como desarrollo de lo anterior, es fundamental cuidar la participación, roles y responsabilidades de los colectivos implicados. Garantizamos la calidad de la evaluación si podemos contar con los informantes clave para cada información y si tenemos claros los diferentes roles que cada uno de los colectivos implicados debe tener en la evaluación. Integrar a todos los participantes relevantes en la planificación y diseño de la evaluación aporta credibilidad al proceso y este es el pilar básico de la utilidad (como señalamos anteriormente, una evaluación no será útil si no es creíble). Pero, además, la

evaluación es creíble y útil cuando tiene consecuencias y estas son coincidentes con las previstas. Si la evaluación aporta información, pero no tiene consecuencias claras para la persona o el colectivo evaluado o afectado por la evaluación, no será ni creíble ni útil. Para determinar estas consecuencias es fundamental definir bien cuáles son los participantes y cuáles deben ser sus roles en el proceso de evaluación.

- Finalmente, va a ser fundamental la elaboración del plan de trabajo para llevar a cabo la evaluación propiamente dicha. Así, el plan de trabajo es válido cuando se garantizan dos condiciones: la fidelidad y la generalidad de la traducción.

1. *Fidelidad de la traducción.* Se trata de asegurar que en el diseño del plan de trabajo se recogen todos los elementos (todas las competencias) que hemos definido como demostrativos de la calidad docente, cuidando así tanto la validez interna como la validez de constructo. Asimismo, deben integrarse todos los factores que condicionan o moderan las posibilidades de actuación docente —variables de entrada y contexto mediato (social) e inmediato (institucional)— y los modos en que se llevan a cabo —procesos personales—. Esto se concreta en las diferentes fases de la evaluación:

a) Recogida de información: determinar bien los momentos y el modo de recogida de información,

sin introducir interrupciones o interferencias que modifiquen la realidad y cotidianidad con que se produce habitualmente la situación educativa que se evalúa. Si, por ejemplo, se quiere evaluar si un profesor o profesora imparte bien sus clases, técnicamente lo ideal sería grabar todas las clases y elegir alguna/s al azar; pero, desde una perspectiva ética, esta propuesta presenta problemas que habitualmente esgrime el profesorado (protección de datos, libertad de cátedra, protección de la imagen del profesor...) y que, sin duda, deben ser atendidos desde la deontología profesional antes de apostar por este tipo de soluciones. En cualquier caso, se trata de un tema que debatir, pues creemos que los derechos de los que reciben la docencia tienen prioridad sobre los que tienen quienes la imparten. Por ello, es necesario abordar las decisiones de este tipo a partir de procesos de mediación.

b) Análisis de la información: elegir los análisis adecuados para dar respuesta al objeto y finalidad de la evaluación, elegir análisis que permitan hacer afirmaciones y que esto se traduzca en el plan de evaluación. Por ejemplo, para comprobar el posible incremento de conocimientos de los estudiantes será necesario un diseño pretest-postest, o para comprender el uso que el profesorado hace

de sistemas de evaluación continua como el portfolio será necesario, además de recoger los propios portfolios, establecer un sistema de registro de narraciones sobre los usos que hacen de ellos y cómo perciben esos usos.

- c) Interpretación de la información, desde la integración contextual: más allá de la definición del perfil docente de calidad para cada nivel educativo y cada modalidad, la evaluación debe también ser sensible a las características del contexto (medio rural o urbano, nivel socioeconómico de las familias de procedencia del alumnado...). Es muy importante para modular la interpretación de resultados, pero esta integración de los elementos contextuales es muy difícil de realizar porque esa modulación de interpretación de resultados solo se puede sustentar en el enfoque cualitativo, en la observación del evaluador, siendo esta una de las evaluaciones más difíciles de realizar.
- d) Uso de la información: extraída y analizada la información, es necesario ajustar el uso a las finalidades manifestadas para la evaluación. Si se diseña la evaluación formativa para la mejora, la información obtenida requiere un análisis más pormenorizado, indicando cuáles son las lagunas o deficiencias identificadas y cuáles son los ele-

mentos que se pueden dinamizar para la mejora.

2. *Generalidad de la traducción.* La calidad del plan de trabajo, como pilar de la validez de la evaluación, depende de que ese plan de trabajo refleje todos los aspectos implicados en el concepto de calidad docente y el propósito y finalidad que animaron el diseño del plan de evaluación para que fuera una herramienta dirigida a conseguir determinadas consecuencias. Evaluaciones parciales como valorar solo el dominio de la materia, las competencias de planificación o el uso de las TIC no son adecuadas; solo si se integran de forma coherente todos los componentes se puede hablar de una evaluación válida, cumpliendo el criterio de generalidad de la traducción del objeto de evaluación, vinculado con la validez de constructo.

Se presentan como ejemplo dos situaciones evaluativas del ámbito de la educación superior en España que probablemente no se hubieran dado de haber respetado estos consejos y, especialmente, de existir un contrato de evaluación. En 1989 se inicia en España la evaluación del profesorado universitario: evaluación de la investigación y de la docencia.

La evaluación de la investigación se diseña en 1989 a nivel estatal, con finalidad formativa y de dotación de un complemento salarial (Tejedor y Jornet, 2008). Se propuso en períodos de seis años, voluntaria y realizada a partir de la valoración de los cinco trabajos seleccionados como más relevantes por el profesor. Así presentada, fue bastante

bien aceptada por el colectivo profesional. Sin embargo, con el transcurso de los años, ha pasado a tener consecuencias sumativas que van más allá del complemento salarial inicialmente acordado. Actualmente, se han derivado de esa evaluación otras muchas consecuencias (participación en tribunales de tesis o de promoción de profesorado, participación en dirección de tesis, consecuencias para la dedicación docente del profesorado) no previstas en el diseño inicial.

La evaluación de la docencia en la universidad española se inició con proyectos puntuales: a) con carácter experimental, dirigidos al diseño de instrumentos y para ir adaptando la realidad de la evaluación al contexto de nuestras universidades (Aparicio, Tejedor, y Sanmartín, 1984), y b) con usos y finalidades institucionales como, por ejemplo, el de la Universitat de València en el curso 1986/87, que estaba esencialmente vinculado a procesos de mejora de la docencia diseñados a partir de los resultados de la evaluación. Esta orientación contribuyó a salvar las grandes reticencias iniciales del profesorado a ser evaluado (y que estas evaluaciones se realizaran tomando las opiniones de estudiantes como fuente de información principal), permitiendo así la implantación del modelo. Sin embargo, en cuanto cambió la coordinación del proceso, perdió esa función formativa, se dejaron de diseñar planes de formación adaptados a los resultados evaluativos y se limitó a un uso sumativo (consecuencias para incremento salarial) que después se ha intensificado, vinculándolo también a procesos de promoción docente.

Estos cambios en los usos de los resultados de la evaluación no son éticos (ni en el

evaluador ni en el legislador). Si hubiera habido un planteamiento profesional o un contrato especificando los usos y consecuencias que tendría la evaluación, no se hubieran podido cambiar las consecuencias. Si los responsables hubieran considerado necesario introducir cambios, hubiera sido requerido un nuevo acuerdo. Por esto es tan importante la participación real de los colectivos en la evaluación, para garantizar la aceptación de la evaluación, que esa aceptación quede por escrito y en qué condiciones, como se da en algunos países. Estos componentes éticos son fundamentales en procesos de evaluación que tienen consecuencias directas o indirectas en el desarrollo profesional de personas concretas. En definitiva, como es evidente, se incumplió la ya mencionada validez consecucional, de forma que en la actualidad este tipo de procesos han quedado como un hecho administrativo carente de sentido para docentes y alumnado, pero que satisfacen la *imagen institucional comprometida con el control evaluativo de su acción*.

#### 4.2. Estrategias de cuidado de la validez de la evaluación a través de la gestión de la información

El segundo conjunto de estrategias que permiten cuidar la validez en los procesos de evaluación de la docencia se refiere a la gestión de la información, y se estructuran en torno a las sucesivas fases de la evaluación.

1. Antes de iniciar la evaluación. La planificación de la evaluación se ha de focalizar en la definición exhaustiva del objeto de evaluación, esto es, el perfil del docente que se quiere evaluar. La definición pormenorizada de cada uno de sus elementos puede

asegurar que vayamos a evaluar los componentes clave para el proceso de toma de decisiones.

2. Durante el proceso de evaluación es necesario cuidar la flexibilidad de la planificación. La planificación previa debe recoger un protocolo, pero también debe prever los elementos de adaptación a la situación, a partir de los elementos facilitadores y obstaculizadores que se hayan podido identificar a priori (Nussbaum, 2006; Walker, 2003). Además, en el proceso de recogida de información, es necesario cuidar la calidad de los instrumentos de medida, de acuerdo con los principios de la medición educativa como ámbito disciplinar. Pese al desarrollo histórico de la evaluación, en algunos contextos se sigue asumiendo medición y evaluación como sinónimos, cuando no lo son. La medición es únicamente un proceso instrumental dentro del proceso de evaluación, pero es un proceso fundamental, que debe velar por la calidad de los instrumentos, tanto de los estandarizados como de las técnicas evaluativas, más de corte cualitativo, también fundamentales en los procesos de evaluación de la docencia.
3. Después de realizar el proceso de evaluación es imprescindible cuidar la difusión y uso de la evaluación, siendo procesos habitualmente desatendidos. Para que la evaluación sea creíble y sea útil debe ser transmitida de forma comprensible, explicando bien sus alcances y sus limitaciones.

Para finalizar, destacamos que, si la evaluación utiliza estas estrategias para cuidar la validez a lo largo de todo el proceso, dando el papel que requiere a cada uno de los colectivos implicados, en función de su objeto y finalidad, entonces la evaluación está asumiendo claramente un rol de mediación en los procesos. Esta lectura de la evaluación como mediación se puede dar a distintos niveles:

- En primer lugar, en los acuerdos previos a la propia evaluación. Siguiendo el concepto de *negociación* de Cronbach (1963), y según se ha recogido en las estrategias anteriores, la primera fase de la evaluación incluye la negociación con los colectivos implicados, para establecer de forma explícita todos los elementos integrados en la propia evaluación, incluyendo el uso y la difusión de resultados.
- La planificación de la evaluación supone un nuevo proceso de mediación, como *compromiso entre lo deseable y lo posible*, entre la definición del objeto de evaluación consolidada en el marco teórico y su traducción posible a la situación natural evaluada, siendo realista con las condiciones de la evaluación, evitando interrupciones y, como se indicaba en las estrategias, flexibilizando el plan de trabajo para adecuarlo a las condiciones reales de la situación evaluada. Esto supondrá también la necesaria modulación de la interpretación de datos y el uso de los mismos, que estará condicionada no solo por lo que preveía el modelo, sino por lo que realmente se ha podido concretar en el proceso de evaluación.



- Específicamente y vinculado con el uso de la evaluación y sus consecuencias, la evaluación, como se decía anteriormente, siempre va a ser un acto de *mediación entre la realidad y la interpretación* que se hace de la misma a partir de los resultados de la evaluación.

## 5. Conclusiones

En este trabajo hemos intentado aportar un marco global que sirva para conceptualizar la validación de procesos de evaluación de la docencia. Trabajar en la validación de la evaluación de la docencia es un ámbito que se está iniciando. Se han realizado aportaciones previas, pero en elementos parciales, como en la validación de los instrumentos de evaluación de la docencia por parte del alumnado. Ahora es importante dedicar esfuerzos para aportar un análisis integrador de la validación de la docencia, tanto en los elementos generales de validación, como en los elementos de especificación que sea necesario establecer. Así, estos elementos de especificación se concretarán para la evaluación en función del nivel educativo y de la modalidad con que se desarrolla la docencia, partiendo en todo caso de una adecuada especificación del perfil docente de calidad y de las competencias docentes esperadas, y desarrollar, paralelamente, estrategias específicas que sirvan de soporte a los equipos que están diseñando e implementando efectivamente las evaluaciones de la docencia, para que estas sean de calidad, algo totalmente imprescindible teniendo, como tienen, consecuencias concretas y, en ocasiones, duras, en el desarrollo profesional de personas, colectivos e instituciones.

No obstante, y tal como se habrá podido apreciar a partir de esta propuesta, el reflexionar sobre procesos de validación en la evaluación de la docencia, implica abrir el debate sobre:

- a) El estatus de la evaluación en relación con la investigación.
- b) La identificación de una posición epistemológica que oriente los procesos evaluativos (de su posicionamiento indudablemente se derivarán también consideraciones relacionadas con los procesos de validación).
- c) Las aplicaciones particulares sobre validación que deberían enfatizarse en función de su aplicación en diferentes tipos de docencia (presencial vs. a distancia, niveles —desde educación infantil a educación superior—).
- d) Las tipologías de evaluación según su finalidad (rendición de cuentas vs. orientada a la mejora), etc.

Se trata, por tanto, de una propuesta primigenia que deberá ser desarrollada y especificada, así como debatida en relación con los enfoques y soluciones metodológicos que puedan ser más apropiados para cada caso.

Este ejercicio de reflexión que compartimos, lo realizamos desde el convencimiento de que validar la evaluación es hacerla útil para los implicados, porque si la evaluación no es útil para nadie, en definitiva, tampoco es útil para nada. La validez es la clave para que los procesos de evaluación tengan sentido y es la base para asegurar la justicia y la equi-

dad en la evaluación de cualquier fenómeno, en general, y de la docencia, en particular. Por desgracia, estamos observando, ya con demasiada frecuencia, evaluaciones que tienen graves problemas de validez (desde las facetas genéricas hasta las específicas) y que, sin embargo, se usan con orgullo por parte de administradores y políticos que justifican con ello que la educación ha entrado en una cultura de evaluación. Desde nuestra posición, trabajar por el desarrollo de una verdadera cultura de evaluación se realiza cuestionando la validez de las evaluaciones, analizando en profundidad sus diseños y aplicación, sus usos y sus consecuencias y, para ello, la base es construir una referencia que nos ayude a conceptualizar la validez de las mismas. Entendemos que este ejercicio es la base del compromiso ético-deontológico que debe presidir cualquier hecho evaluativo. Por ello, animamos a los investigadores en educación y, en especial, a los especialistas en medición y evaluación educativas a sumarse a este compromiso para desarrollar un corpus conceptual que oriente la calidad de la evaluación de la docencia.

## Notas

<sup>1</sup> Se puede entender estableciendo un paralelismo cognitivo con lo que se produce en medición en la construcción de instrumentos estandarizados de medida, donde la fiabilidad es una condición necesaria, pero no es suficiente para asegurar la validez.

<sup>2</sup> En muchos de los procesos administrativos desarrollados por las Administraciones Públicas se utilizan modelos que se asume que, por genéricos, son justos y equitativos, sin percatarse de su falta de validez. Esta carencia se da, además, de forma diferencial entre los diversos centros al no reflejar en el modelo de evaluación las singularidades institucionales.

<sup>3</sup> Tradicionalmente, se menciona la negociación como una etapa fundamental para ajustar el plan de evaluación. Desde nuestro punto de vista, entendemos que

se trata más bien de un proceso de mediación que de negociación. El objetivo no es que ninguna parte pierda para poder llevar a cabo un proceso, sino que todas las partes o colectivos implicados en la evaluación se entiendan y conseguir que la evaluación aporte elementos de mejora. Debe asumirse como un proceso que favorece a todos.

## Referencias bibliográficas

- Anguera, M. T., Blanco, A., Losada, J. L. y Portell, M. (2018). Pautas para elaborar trabajos que utilizan la metodología observacional. *Anuario de psicología*, 48 (1), 9-17.
- Aparicio, J. J., Tejedor, F. J. y Sanmartín, J. (1984). Reseñas de investigación. La enseñanza universitaria vista por los alumnos: un estudio para la evaluación de los cursos en la enseñanza superior. *Revista de Educación*, 273, 238-239.
- Bericat Alastuey, E. (1998). *La integración de los métodos cuantitativo y cualitativo en la investigación social: significado y medida*. Madrid: Ariel.
- Cronbach, L. J. (1963). Course improvement through evaluation. *Teachers College Record*, 64 (8), 672-683.
- Delgado Álvarez, C. (2014). *Viajando a Ítaca por mares cuantitativos: manual de ruta para investigar en grado y postgrado*. Salamanca: Amarú.
- García Mestanza, J. (2010). Propuesta de evaluación de la actividad docente universitaria en entornos virtuales de aprendizaje. **revista española de pedagogía**, 68 (246), 261-280.
- González-Sanmamed, M., Muñoz-Carril, P. C. y Santos-Caamaño, F. J. (2019). Key components of learning ecologies: A Delphi assessment. *British Journal of Educational Technology*, 50 (4), 1639-1655.
- González-Sanmamed, M., Sangrá, A., Souto-Seijo, A. y Estévez, I. (2018). Ecologías de aprendizaje en la era digital: Desafíos para la educación superior. *Publicaciones*, 48 (1), 11-38.
- González-Such, J. (1988). *Evaluación de la docencia universitaria*. Valencia: CSV.
- Guerra García, M., Leyva Barajas, Y. E. y Conzuelo Serrato, S. (2019). Factores contextuales que afectan la práctica docente desde la perspectiva de docentes en México. *Publicaciones*, 49 (1), 137-149. doi:10.30827/publicaciones.v49i1.9858

- Jornet Meliá, J. M. (septiembre, 2014). Facetas del concepto de validez en la evaluación de la Docencia. En *V Coloquio de la Red Iberoamericana de Investigadores sobre Evaluación de la Docencia (RIIED)*. Simposio llevado a cabo en el congreso del Instituto de Investigación y Desarrollo Educativo de la Universidad Autónoma de Baja California (UABC), Ensenada, México.
- Jornet Meliá J. M., Bakieva, M. y Sánchez-Delgado, P. (en prensa). La cohesión social como objetivo de la educación: ¿Podemos especificar un modelo de calidad para realizar la evaluación de sistemas educativos? *Fronteiras: Journal of Social, Technological and Environmental Science*.
- Jornet Meliá, J. M., González-Such, J., Perales Montolío, M. J., Sánchez-Delgado, P., Bisquert, M., Bakieva, M., ... Ortega Gaite, S. (2017). Aproximaciones cualitativas para la definición y validación de constructos de instrumentos estandarizados de medida. En A. P. Costa, M. C. Sánchez-Gómez y V. Martín Cillero (Eds.), *La Práctica de la investigación cualitativa: ejemplificación de estudios*. Salamanca: CIAIQ.
- Jornet Meliá, J. M., Suárez Rodríguez, J. M. y Perales Montolío, M. J. (2003). *Metodología de evaluación de programas de formación ocupacional y continua*. Valencia: Aedit.
- Jornet Meliá, J. M., Suárez Rodríguez, J. M. y Pérez Carbonell, A. (2000). La validez en evaluación de programas. *Revista de Investigación Educativa*, 18 (2), 341-356.
- Lorenzo Quiles, O. (2011). Análisis cualitativo de textos sobre multi e interculturalidad. *DEDi-CA. Revista de Educação e Humanidades*, 1, 535-546.
- Manzi, J., González, R. y Sun, Y. (Eds.) (2011). *La evaluación docente en Chile*. Santiago de Chile: MIDE-UC.
- Martínez-Rizo, F. (Coord.). (2015). *Las pruebas ENLACE y EXCALE. Un estudio de validación. Cuadernos de Investigación*, 40. México, DF: Instituto Nacional para la Evaluación de la Educación.
- Martín-García, A. V., Sánchez-Gómez, M. C. y Costa, P. (2019). Percepción de Blended Learning en profesores universitarios de distintos ámbitos disciplinares. *Revista Lusófona de Educação*, 44 (44), 117-133.
- Mertens, D. M. (2015). Mixed methods and wicked problems. *Journal of Mixed Methods Research*, 9 (1), 3-6.
- Nava, M. y Rueda, M. (2014). La evaluación docente en la agenda pública. *Revista Electrónica de Investigación Educativa*, 16 (1), 1-11.
- Nussbaum, M. C. (2006). Education and democratic citizenship: capabilities and quality education. *Journal of Human Development*, 7 (3), 385-395.
- Ortega Gaite, S., Tejedor Mardomingo, M. y Perales Montolío, M. J. (septiembre, 2019). La distribución económica y desigualdad en los manuales escolares de educación para la ciudadanía y derechos humanos en clave de educación para el desarrollo. En *Educación Crítica: hacia una práctica inclusiva y comprometida socialmente (EDUCRÍTICA 2019)*. Simposio presentado en el congreso internacional de la Facultad de Educación de la Universidad de León (ULE), León, España.
- Rueda-Beltrán, M. (Coord.) (2019). *Las condiciones institucionales de la docencia en universidades iberoamericanas*. Temuco: Red Iberoamericana de Investigadores sobre Evaluación de la Docencia (RIIED), Universidad de La Frontera.
- Ruiz-Primo, M. A., Jornet Meliá, J. M., Backhoff Escudero, E. (2006). *Acerca de la Validez de los Exámenes de la Calidad y el Logro Educativos (Excale). Cuadernos de Investigación (México)*, 20. Ciudad de México: Instituto Nacional para la Evaluación de la Educación. Recuperado de <https://www.inee.edu.mx/publicaciones/acerca-de-la-validez-de-los-examenes-de-la-calidad-y-el-logro-educativas-excale-no-20/> (Consultado el 25-08-2019).
- Rutman, L. (1984). *Evaluation research methods: a basic guide*. California: Sage Publications.
- Sánchez-Gómez, M. C., Martín-Cilleros, M. V., Costa, A. P. y Peñalvo, F. J. G. (2018). Posicionamiento de la investigación en Ciencias Sociales. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, 28, 102-113.
- Sánchez-Gómez, M. C., Rodrigues, A. I. y Costa, A. P. (2018). Desde los métodos cualitativos hacia los modelos mixtos: tendencia actual de investigación en ciencias sociales. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, 28.
- Sandín Esteban, M. P. (2000). Criterios de validez en la investigación cualitativa. De la objetividad a la solidaridad. *Revista de Investigación Educativa*, 18 (1), 223-242.

- Solano-Flores, G., Contreras-Niño, L. A. y Backhoff, E. (2006). Traducción y adaptación de pruebas: Lecciones aprendidas y recomendaciones para países participantes en TIMSS, PISA y otras comparaciones internacionales. *Revista Electrónica de Investigación Educativa*, 8 (2).
- Solano-Flores, G. y Milbourn, T. (2016). Capacidad evaluativa, validez cultural y validez consecuen- cial en PISA. *RELIEVE Revista Electrónica de Investigación y Evaluación Educativa*, 22 (1).
- Taut, S. y Palacios, D. (2016). Interpretaciones no intencionadas e intencionadas y usos de los resultados de PISA: Una perspectiva de validez consecuen- cial. *RELIEVE Revista Electrónica de Investigación y Evaluación Educativa*, 22 (1).
- Tejedor, F. J. (2004). Investigación educativa: ¿Hacia dónde vamos? En L. Buendía, D. González y T. Pozo (Coords.), *Temas fundamentales en la investi- gación educativa* (pp. 63-107). Madrid: La Muralla.
- Tejedor, F. J. y García Valcárcel, A. (2010). Evalua- ción del desempeño docente. *revista española de pedagogía*, 68 (247), 439-459.
- Tejedor, F. J. y Jornet Meliá, J. M. (2008). La eva- luación del profesorado universitario en Espa- ña. *REDIE Revista Electrónica de Investigación Educativa*, 10, número especial 2008, 1-29.
- Vedung, E. (1995). Utilización de la evaluación de programas. En R. Fernández Ballesteros (Ed.), *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud* (pp. 321-338). Madrid: Síntesis
- Walker, M. C. (2003). Framing social justice in ed- ucation: what does the «capabilities» approach offer? *British Journal of Educational Studies*, 51, 168-187.
- Weiss, C. (1998). *Evaluación*. Nueva Jersey: Pren- tice-Hall.

## Biografía de los autores

**Jesús Miguel Jornet Meliá** es Doc- tor en Filosofía y Ciencias de la Educación/ Pedagogía por la Universitat de València y Catedrático de Medición y Evaluación Edu- cativas. Fue Director del Departamento de Métodos de Investigación y Diagnóstico en Educación (MIDE) entre 1993 y 1999 en di-

cha universidad. Actualmente, coordinada el grupo GemEduCo: [www.uv.es/gem](http://www.uv.es/gem). Líneas de investigación: diseño de instrumentos de medición y evaluación, evaluación de competencias, programas, profesorado y sistemas educativos.

 <https://orcid.org/0000-0001-6905-497X>

**María Jesús Perales Montolío** es Doc- tora en Filosofía y Ciencias de la Educación por la Universitat de València. Actualmente, es Profesora Titular del Departamento de Métodos de Investigación y Diagnóstico en Educación; Directora del Máster Oficial en Acción Social y Educativa, y miembro del Grupo de investigación GemEduCo. Líneas de investigación: diseño de instrumentos de medición y evaluación, evaluación de compe- tencias, programas, profesorado y sistemas educativos.

 <https://orcid.org/0000-0003-2033-2750>

**José González Such** es Doctor en Fi- losofía y Ciencias de la Educación. Actual- mente, es Catedrático del área de Métodos de Investigación y Diagnóstico en Educación y Director del Máster de Educación Especial de la Universitat de València; Miembro del grupo GemEduco, y Coordinador de la Red de Innovación educativa y Calidad Docen- te InnovaMIDE. Líneas de investigación: diseño de instrumentos de medición y eva- luación, evaluación de competencias, pro- gramas, profesorado y sistemas educativos; innovación educativa.

 <https://orcid.org/0000-0001-9086-6446>

# Sumario\*

## Table of Contents\*\*

### Editorial Editorial

---

**José Antonio Ibáñez-Martín**

Una consideración educativa sobre la pandemia:  
resistir... y adelantar

*An educational consideration on the pandemic: endure...  
and progress*

181

### Estudios Studies

---

**Bernardo Gargallo López, Fran J. García-García,  
Inmaculada López-Francés,  
Miguel Ángel Jiménez Rodríguez y  
Salomé Moreno Navarro**

La competencia aprender a aprender:  
valoración de un modelo teórico

*The learning to learn competence: An assessment of a  
theoretical model*

187

**Miguel A. Santos Rego, María José Ferraces Otero,  
Ígor Mella Núñez y Ana Vázquez-Rodríguez**  
Universidad, competencias cívico-sociales  
y mercado de trabajo

*University, civic-social competences,  
and the labour market*

213

**Jesús Miguel Jornet Meliá, María Jesús Perales  
Montolio y José González-Such**

El concepto de validez de los procesos  
de evaluación de la docencia

*The concept of validity of teaching evaluation processes* 233

**Fernando Acevedo Calamet**

Factores explicativos del abandono de los estudios  
en la educación superior en contextos socio-  
académicos desfavorables

*Explanatory factors for dropout from higher education in  
unfavourable socio-academic contexts*

253

### Notas Notes

---

**María-Carmen Ricoy y Cristina Sánchez-Martínez**

Revisión sistemática sobre el uso de la tableta en la  
etapa de educación primaria

*A systematic review of tablet use in primary education* 273

**Jaime Carcamo-Oyarzun y Christian Herrmann**

Validez de constructo de la batería MOBAC para la  
evaluación de las competencias motrices básicas en  
escolares de educación primaria

*Construct validity of the MOBAC test battery for the assessment  
of basic motor competencies in primary school childrens* 291

\* Todos los artículos están también publicados en inglés en la página web de la revista: <https://revistadepedagogia.org>.

\*\* All the articles are also published in English on the web page of the journal: <https://revistadepedagogia.org>.

**Concha Iriarte Redín, Sara Ibarrola-García y Maite Aznárez-Sanado**

Propuesta de un instrumento de evaluación de la mediación escolar (CEM)

*Proposal for a school mediation evaluation tool (MEQ)* 309

**Bruno Echauri Galván y Silvia García Hernández**

Traducir en colores: la traducción como herramienta de evaluación de la comprensión lectora en inglés como lengua extranjera

*Translating in colours: Translation as an assessment tool for reading comprehension in English as a second language* 327

## **Reseñas bibliográficas**

---

**Esteban-Bara, F. (2019).** *La universidad light: Un análisis de nuestra formación universitaria* (José L. González-Geraldo).

**Ruiz-Corbella, M. y García-Gutiérrez, J. (Eds.) (2019).** *Aprendizaje-Servicio. Los*

*retos de la evaluación* (Andrea Muñoz Villanueva).

**Lafforgue, L. (2019).** *Recuperemos la escuela* (Beatriz Gálvez). **Ferraces-Otero, M. J., Godás-Otero, A. y García-Álvarez, J. (2019).** *Cómo realizar un estudio científico en ciencias sociales, de la educación y de la salud* (Carolina Rodríguez-Llorente). 347

## **Informaciones**

---

46.ª conferencia anual de la Association for Moral Education sobre el tema: «Morality, Environmental Sustainability, and Education»; XV Congreso Internacional de Teoría de la Educación: «Democracia y tradición en la teoría y la práctica educativa del siglo XXI. En el 50 aniversario de la Ley General de Educación». 361

## **Instrucciones para los autores**

Instructions for authors 365



ISSN: 0034-9461 (Impreso), 2174-0909 (Online)

<https://revistadepedagogia.org/>

Depósito legal: M. 6.020 - 1958

INDUSTRIA GRÁFICA ANZOS, S.L. Fuenlabrada - Madrid

# The concept of validity of teaching evaluation processes

## El concepto de validez de los procesos de evaluación de la docencia

Jesús Miguel JORNET MELIÁ, PhD. Professor. Universitat de València ([jornet@uv.es](mailto:jornet@uv.es)).

María Jesús PERALES MONTOLÍO, PhD. Reader. Universitat de València ([perales@uv.es](mailto:perales@uv.es)).

José GONZÁLEZ-SUCH, PhD. Professor. Universitat de València ([jose.gonzalez@uv.es](mailto:jose.gonzalez@uv.es)).

### Abstract:

This paper presents a conceptual and methodological proposal about validity in teaching assessment processes. Although they are very common processes, and in some cases have significant individual and organisational consequences, a theoretical and methodological approach to them has not yet been performed. Based on reviewing research and the experience accumulated over years of conducting evaluations and evaluative research, we take as a starting point our own conceptualisation of evaluation (summative and formative evaluation or appraisal/diagnosis) as an applied research form that essentially uses measurement processes. Therefore, it takes the concept of validity in these two knowledge areas and

applies it to the area of teaching evaluation by providing a specific definition of the validity in the evaluation process. Moreover, based on conceptual references in research and measurement, this work identifies generic and specific methodological considerations, which include the concept of validity, providing meaningful definitions for each of them. Understanding that validity, as a quality of evaluation, must be taken into account throughout the evaluation process, we also propose a series of strategies to ensure this is done correctly, based on defining validity and the aspects that comprise it. Therefore, this work's contribution is conceptual and methodological, and it proposes a reference point for conceptualising validity in a field with broad use and with profound conse-

Revision accepted: 2019-12-21.

This is the English version of an article originally printed in Spanish in issue 276 of the **revista española de pedagogía**. For this reason, the abbreviation EV has been added to the page numbers. Please, cite this article as follows: Jornet Meliá, J. M., Perales Montolío, M. J., & González-Such, J. (2020). El concepto de validez de los procesos de evaluación de la docencia | *The concept of validity of teaching evaluation processes*. *Revista Española de Pedagogía*, 78 (276), 233-252. doi: <https://doi.org/10.22550/REP78-2-2020-01>  
<https://revistadepedagogia.org/>

ISSN: 0034-9461 (Print), 2174-0909 (Online)

revista española de pedagogía  
year 78, n. 276, May-August 2020, 233-252



quences for teachers, institutions, and educational systems.

**Keywords:** evaluation, faculty, validity, research, measurement.

## Resumen:

En este trabajo se presenta una propuesta conceptual y metodológica sobre la validez en los procesos de evaluación de la docencia. Pese a ser procesos muy habituales, y en algunos casos con significativas consecuencias individuales y organizacionales, no se ha realizado todavía una aproximación teórica y metodológica sobre la validez de los mismos, siendo este el objetivo de este trabajo. A partir de la revisión bibliográfica y de la experiencia acumulada durante años realizando evaluación e investigación evaluativa, tomamos como punto de partida una conceptualización propia de la evaluación (sumativa y formativa o valoración/diagnóstico) como una modalidad de investigación aplicada, que utiliza de forma esencial procesos de medición. Por ello, toma

el concepto de validez en estos dos ámbitos de conocimiento y los lleva al área de la evaluación de la docencia, aportando una definición específica de validez de la evaluación de la docencia. Además, a partir de los referentes conceptuales en investigación y en medición, concreta las facetas metodológicas, genéricas y específicas que integran ese concepto de validez, aportando definiciones propias para cada una de ellas. Finalmente, entendiendo que la validez, como calidad de la evaluación, debe ser cuidada durante todo el proceso evaluativo, partimos de su definición y de las facetas que la constituyen y proponemos una serie de estrategias para poder asegurarla de forma adecuada. Por lo tanto, la aportación de este trabajo es de carácter conceptual y metodológico, y propone un punto de referencia para conceptualizar la validez en un ámbito de amplio uso y con profundas consecuencias para los profesores, las instituciones y los sistemas educativos.

**Descriptor:** evaluación, profesorado, validez, investigación, medición.

## 1. Introduction

Teaching evaluation has been gradually implemented to support managing, innovation and improving education quality. It is in higher education where the most experiments have been done with consequences for accountability (García Mestanza, 2010; Tejedor & García-Valcárcel, 2010). At pre-university levels, these objectives have had a stronger influence in recent years (Guerra García, Leyva Ba-

rajas, & Conzuelo Serrato, 2019; Manzi, González, & Sun, 2011; Nava & Rueda, 2014). Nonetheless, some major gaps in studies about the validity of evaluations appear when reviewing the literature on the evaluation methodology of teaching or teachers.

In this work, we examine some methodological reflections that help reconsider aspects involved in the validity of teaching



evaluation. Before going into greater depth with this analysis, we believe it is necessary to make some observations. Firstly, as regards the use of the terms *teaching evaluation* or *teacher evaluation*, we understand that approaches sometimes differ. Use of the term *teaching* is preferred in approaches that assume that the teacher's activity occurs in an institutional context, which basically means that what is individual cannot be analysed independently of the context in which the teacher works. However, when the term *teacher evaluation* is used, emphasis is placed on the evaluation of professional work. We take both to be synonyms because this article is about validity. We cannot understand teachers' professional evaluation outside the context of the institutional setting in which teachers work.

Secondly, we note that this point of view does not deny that the different categories of teacher (according to level of education, subject, etc.) and the contexts in which they work (face-to-face/distance teaching, public/private institutions, the contexts in which teaching takes place —vulnerable vs. overstimulated, etc.) require a specific analysis of the factors involved in the validity of their evaluation, because if we seek quality evaluations in each particular case, then evaluation processes need to recognise these peculiarities. However, in this work, we take a general approach at the possible expense of neglecting certain specific points, which we hope to examine in subsequent analyses of particular teaching typologies.

This review is arranged into the following sections: the concept of validity in evaluation processes, aspects of the concept of validity, and strategies for achieving validity.

## 2. On the concept of validity

The concept of validity has been the focus of attention as a criterion of goodness both in basic research processes in social and educational sciences, and also for designing measurement and evaluation instruments. So, in any area that affects measurement and evaluation in educational sciences, validity is a key theme, not only for building scientific knowledge, but also particularly, in relation to the adequacy of decision making and suitability of intervention proposals.

The concept of validity is multidimensional and heterogenous because it can take different forms when applied to different units or objects of educational research, measurement or evaluation. Validation requirements are far from being uniform and depend, to a great extent, on adjusting the process as a whole to the object's requirements, purposes and characteristics. Therefore, an approach to the topic in the context of evaluation in education requires a certain adaptation of concepts, bearing in mind that, although we are able to identify the common elements to be applied to any given situation, paying attention to the specific characteristics each has might also imply placing a different emphasis on the particular meanings or approaches involved in validity.

We begin with a concept of validity in evaluation that we have been developing for several years, which we initially formulated to evaluate programmes as a discipline (González-Such, 1988; Jornet Meliá, Suárez Rodríguez, & Pérez Carbonell, 2000; Ruiz-Primo, Jornet Meliá, & Backhoff, 2006; Jornet meliá et al., 2017), and for teaching evaluation (Jornet Meliá, 2014). This concept comes from the experience accumulated from several research projects (UNIVECS [EDU2016-78065-R], SECS/EVALNEC [EDU2012-34734], EVALEF [EDU2011-29467], MAVACO [EDU2009-13485], and AVACO [SEJ 2005-05 923]), developed by our research group GemEduco (<https://www.uv.es/gem/gemeduco/>) and others (Rueda-Beltrán, 2019), in collaboration with RIIED, the Ibero-American Network of Teaching Evaluation Researchers (<http://iide.ens.uabc.mx/ried/ibero/>).

Although attempts have been made to transfer the concept of validity from research and measurement to evaluation, we believe that we have still not been able to provide an adaptation that satisfies the needs to conceptualise the quality — or validity — of evaluation.

Post-positivist evaluation positions provide the concepts of validity that are used in measurement and in basic research. If we assume that evaluation is an applied form of research, which still has conceptual difficulties, the change in how validity is understood could be partially implemented (Delgado Álvarez, 2014). We are aware that this position, taking evaluation as a form of applied research, is not shared by all authors. Nonetheless, we understand

evaluation as a form of applied research in education insofar as evaluation:

- a) Provides information about the phenomenon of education in natural contexts without manipulating variables.
- b) Makes it possible to identify and distinguish the characteristics linked to satisfactory and unsatisfactory practices, and also to factors that can be revitalised to improve education in a given context.
- c) Evaluation, if it is seen as a process to understand the educational reality, at least has characteristics of an «inductive view of the educational process». This means that it can identify elements that help generate research questions, considering objectives and hypotheses, and, consequently, proposing possible theories emanating from direct analysis of educational reality.

Furthermore, from a qualitative evaluation perspective, the concept of validity takes shape in approach strategies, such as using multiple sources and instruments and triangulating data (Sandín Esteban, 2000).

In both post-positivist and qualitative approaches, the underlying logics coincide in a good number of cases. For instance, triangulation is a similar logical reasoning strategy to that of the criterion validation of instruments. In the case of triangulation, the difference with the criterion validation of instruments

is that it is not assumed that a criterion actually exists against which to judge the quality of the data collected by the evaluation process being validated, but rather three or more sources or instruments are compared to reach an intersubjective consensus (Jornet Meliá et al., 2017; Jornet Meliá, Bakieva, & Sánchez-Delgado, in press).

In this way, considering that references are scarce and normally far removed from specific teaching evaluation cases, it is not easy to adapt the concept of validity well for this area. Therefore, we direct our attention towards adapting the concept of validity of teaching evaluation starting from two basic benchmarks: the usual concepts of validity in basic research and the different meanings of validity employed in measurements.

We now review the elements that condition the concept of validity:

1. *The concept of evaluation.* This is a systematic data collection process that is directed at enquiry in order to *understand* reality so that we can use these data as a basis to make decisions that help to improve educational interventions.
2. *Teaching work is systemic action.* Teaching evaluation goes beyond understanding the specific intervention that teachers might perform in the teaching-learning process by assuming that teaching work is a global action. This depends on the context in which

teachers work as their tasks can vary greatly and should, in all cases, respond to characteristics of culture and institutional reality. To this end, one basic component of teaching evaluation would be to design an evaluation model that adapts to institutional reality. *The evaluation must be done in natural settings.* Unlike in basic research, we cannot manipulate reality to ensure the quality of the interpretation of results (Tejedor, 2004; Sánchez-Gómez, Martín-Cilleros, Costa, & Peñalvo, 2018). This is so because evaluation is done in real contexts, making efforts to ensure that the evaluation does not introduce any factors that change reality. In this case, validating the evaluation is similar to validating a natural observation (Anguera, Blanco, Losada, & Portell, 2018). Consequently, a basic condition for validating the evaluation would be to collect the necessary data in a real situation by including materials from the work being conducted or used in teaching (Lorenzo, 2011; Ortega Gaité, Tejedor Mardomingo, & Perales Montolío, 2019) without changing the situation we wish to evaluate.

3. *Reality is dynamic and changing.* So, when we evaluate educational processes in general and teaching processes in particular, we must bear in mind that the evaluation must include elements in its planning that make it easy to adapt

and flexible in order to be able to understand any changes that may take place throughout the process. As a result, validity should be an aspect that is contemplated during the whole evaluation process. Therefore, the factors involved in validating the design, its implementation, the results and their interpretation, and how data are used can be analysed. Nonetheless, in this work, we refer to the overall process.

4. *Epistemological guidance and the evaluator's role must be subject to the characteristics of the situation to be evaluated.* To this end, our evaluation position, which will also condition the concept of validity, entails methodological complementarity or mixed quantitative/qualitative models (Bericat Alastuey, 1998; Mertens, 2018; Sánchez-Gómez, Rodrigues, & Costa, 2018).
5. *The evaluation must be useful.* An evaluation makes sense when it is taken as either an education management guide or an element to improve any of the processes involved (teaching-learning, educational, guidance, etc.). We can only consider the evaluation to be valid if the data collected from it are useful for enabling improvements and/or innovation in education. Therefore, the peculiarities of each evaluation type (regardless of whether it is/has a formative/summative
- purpose and/or the evaluation object, such as teaching format, context, etc.) must be respected in evaluation processes (Martín-García, Sánchez-Gómez, & Costa, 2019; González-Sanmamed, Muñoz-Carril, & Santos-Caamaño, 2019; González-Sanmamed, Sangrá, Souto-Seijo, & Estévez, 2018). The data collected for the whole evaluation process must be related to the type of decisions that are based on it.
6. *The evaluation must be done with the same requirements for rigour expected in measurements or in basic research.* Evaluating requires working on reality without modifying it. Its advantage over basic research is that it does not set out to obtain results whose interpretations are generalisable for building scientific knowledge (Tejedor, 2004). Its disadvantage, however, is that it cannot manipulate reality to ensure that the interpretation of the results is beyond doubt.

In summary, considering the elements listed above, we can define the concept of validity in teaching evaluation as the degree of representativeness and usefulness that allow us to interpret the evaluation's outcomes assuming that all elements of the reference teaching profile have been taken into account in the situation or context in which teaching takes place, with no interference with the reality of the educational process we are evaluating.

In short, it is a matter of ensuring the quality of the education evaluation process by checking to see if the methodological components involved allow us to:

- Realistically represent the characteristics of the unit of analysis being addressed.
- Facilitate achievement of the objectives set for the educational evaluation process.
- Respect the interests of all those involved in, or affected by, the evaluation process.
- Represent the context in implementation of the evaluation process, modulating the interpretation of

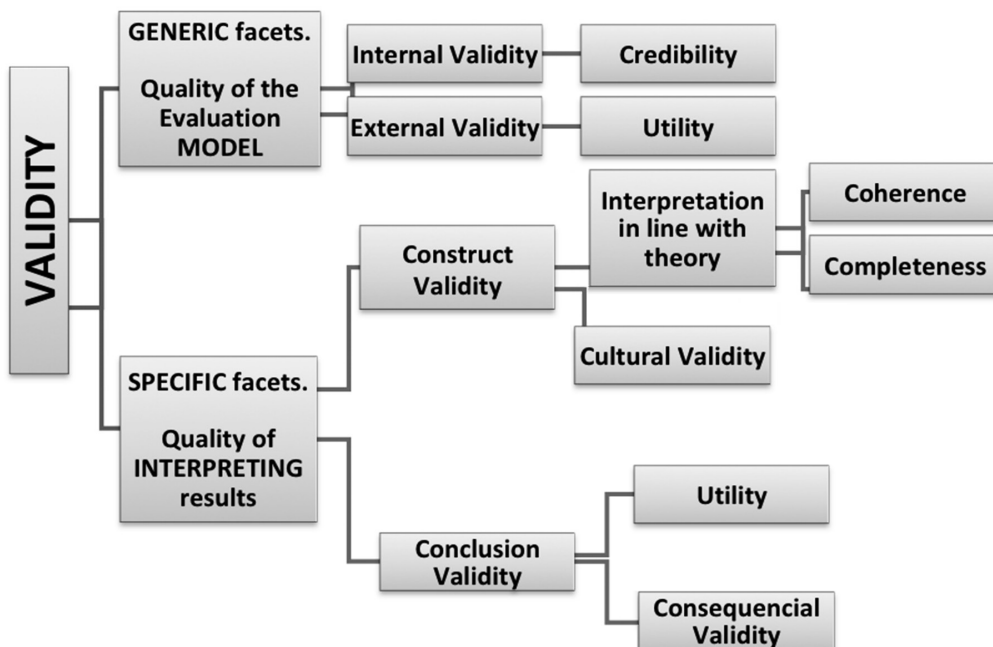
results and in adapting their consequences.

- Contribute to improvement or innovation in the unit under evaluation.

### 3. Facets of the concept of validity

We employ the term *facets* in relation to the different meanings and/or approaches that can be identified in the concept of validity, and we call them methodological because they stem from the classic methodological components of the quality of educational research and measurements. Among them, we can differentiate between topics (Rutman, 1984) that come from research (generic facets), and those that originate from measurements (specific facets) — see Graph 1 —.

GRAPH 1. Classification of concept of validity facets in evaluation.



Source: Own elaboration.

*Generic validity facets* come from post-positivist positions in research and are linked with the quality of the evaluation *model*. As they are applied to define meanings of validity in evaluation, they are present in any process, although their relevance depends on the characteristics of the evaluation's object unit and the evaluation context. Among them, we differentiate between internal and external validity.

By *Internal validity*, we refer to the set of quality criteria that are guided by the principles of *consistency* of the various elements making up the evaluation process. A good level of internal validity for a teaching evaluation process includes the following requirements:

- a) *Internal coherence* among the base indicators to inform the whole evaluation process and subsequent decision making
- b) *Thoroughness* in selecting the criteria and indicators used in the evaluation, i.e., when the evaluation meets the generality characteristic of the object's representation
- c) *Adapting* the type of information included in the evaluation to achieve its *purpose*.

We believe that internal validity is the basis of an evaluation's *credibility*. An evaluation cannot be credible if it does not meet the characteristics set out above. For example, many teaching evaluation systems set out to evaluate teach-

ers by taking their students' outcomes as the only reference. Evidently, this type of evaluation does not fulfil the three criteria listed above. Hence these evaluations are forcefully rejected as they lack credibility. Conversely, if the evaluation process for teaching is based on a thorough, coherent structure of indicators that represent the teacher profile that is considered suitable for the context in which teaching takes place, and it allows the questions of why and what for of the evaluation to be answered, then it is much more likely that this evaluation will be credible for the involved groups and will have internal validity.

Moreover, we believe that an evaluation possesses *external validity* when the information collected and the way in which the whole evaluative process has been directed can be employed for the evaluation's expected use. That is, it is a matter of ensuring that the whole evaluation process, from its design to the collection of data, including how it is analysed and how it is transmitted to the involved groups, is suitable for the *utility* principle that initially inspired the evaluation. Accordingly, we understand that external validity is, to a great extent, the result of internal validity; that is, a good level of external validity in the evaluation cannot be achieved if special care of its internal validity is not taken. Hence, it is not possible to believe that the evaluation is useful if its credibility is not ensured beforehand. Nonetheless, ensuring internal validity is a necessary condition, but it does not suffice to ensure external validity<sup>1</sup>. For example, a

situation might arise in which a teacher evaluation process possesses a good level of internal validity for a formative evaluation. However, if it starts being used — just as it was designed — for summative uses, it completely loses its external validity. This is often the case.

Basically, internal and external validity can be taken as references for the teaching evaluation's credibility and utility, which jointly lead to a virtuous circle in which each one feeds back into the suitability of the other: if an evaluation is not credible, it cannot be useful, but the more useful it is, the more credible it becomes. Internal validity, understood as the evaluation model's robustness, and the basis of an evaluation's credibility, greatly depends on both the evaluator and the designed evaluation process. However, with external validity, understood as the evaluation's utility linked with the evaluation's uses, many spurious factors that introduce biases not foreseen by the evaluator can have a strong influence (Vedung, 1995; Weiss, 1998).

As previously indicated, we distinguish between generic and specific facets when defining the methodological facets of validity. We refer to *specific facets* as a reinterpretation to evaluate the approaches to validity that derive from measurements. We also refer to the quality of *interpreting* results when evaluating teaching (see Graph 1). We include both construct validity and conclusion validity.

*Construct validity* refers to the extent to which all the indicators included in the evaluation plan respond to the theoretical approach that guides the definition of quality we wish to evaluate so that they allow the interpretation of scores aligned with theory. As with standard measuring instruments, content validity is integrated, in that the selected indicators must allow a *coherent and thorough* vision of the phenomenon as a whole to be evaluated. For instance, let us assume that we wish to evaluate the teaching in a Montessori school. Evidently, the teaching profile will be restricted by the methods imposed by the Montessori approach, and it will include all the elements that form part of the teaching activity so that the teacher's duties are appropriate in the school. Therefore, adopting an evaluation approach from the European Foundation for Quality Management (EFQM), or from any model that the education administration deems useful within its geopolitical demarcation, will be to no avail no matter how interesting it is<sup>2</sup>.

Similarly, we include *cultural validity* as a dimension of construct validity. Given that teaching, and any part of the educational process, takes place in a given socio-economic and cultural context, the choice of indicators for the evaluation, and putting them into operation in measurement instruments or evaluative data collection techniques, must adapt well to the context. In social and educational sciences, theorising derives from a context and is not, therefore, always generalisable. For instance, an evaluation model

of primary schools designed according to the EFQM Model might include a teaching profile that includes teaching-learning, innovation and management tasks. Management tasks can include competences related to collaborative teaching work, teamwork, or academic leadership. It would be impossible to apply this type of model in rural schools with students of many levels taught by only one teacher, where the teaching profile should include other competences, among which personal autonomy and paying attention to many facts, which in this case will depend on only one teacher, predominate. Other areas relate to linguistically and culturally adapting instruments (Solano-Flores, Contreras-Niño, & Backhoff, 2006; Solano-Flores & Milbourn, 2016), something which is often not considered and so the instruments are used inappropriately. We must bear in mind that maladjustments are sources of bias and, consequently, affect the interpretation of results directly in the form of lack of validity.

The second specific facet is *conclusion validity*. This refers to the matter of whether the *conclusions* that can be reached through the evaluation *serve the purpose* and objective that justified putting it into practice. In other words, it is a matter of ensuring that the interpretation of the results can respond to the reasons and purposes that justify using the evaluation, whether it is used as an instrument for improvement or for educational accountability.

From the field of research: consistency, generalisability, relevance, and utility are characteristics of conclusion validity. We understand that they are intrinsically involved, but it is necessary to emphasise that they can be combined into two large categories that have developed in the area of evaluation in particular: utility and consequential validity. Utility in this case is a pragmatic reference to external validity; that is, does interpreting the results of the evaluation fulfil the purpose of the evaluation? This is why it should be adapted to the specific teaching typology, context, or situation, and to the purpose of the evaluation (formative/summative) to which the evaluation whose validity we analyse refers.

Finally, and complementing utility in conclusion validity, *consequential validity* is a category that has abruptly entered the area of evaluation with great impetus (Martínez Rizo, 2015; Taut & Palacios, 2016). It essentially refers to whether carrying out an evaluation actually has certain wanted or unwanted consequences. Poor alignment with regards to achieving the outcomes of any evaluation can happen by lack or by excess. For example, general evaluation of primary and secondary education schools in a given region could be carried out to improve teaching in all of them. It could even be explained as an attempt to develop an evaluation culture to make improvements, and so a type of evaluation is designed to help identify good educational practices, which will lead to improved teaching regionally. Apart from the eval-



uation plan considering variables related to the socio-economic and cultural levels of the families using each school, others related to school whether the schools are state owned or privately owned, their size, etc., could be considered. Moreover, the level of student achievement in each school as indicated by standardised tests in various subjects could be taken as a key indicator on which to base the quality evaluation. The data collected would be provided to each school, along with the overall mean of equivalent centres to support self-assessment by the schools themselves. Not diffusing the results could also be ensured. As proposed, this evaluation plan is evidently deficient because it would not include enough input and context variables, and no process variables, so its internal validity would be deficient. As a result of this, there would be shortcomings in its external validity and utility. However, its consequential validity would suffice because it can partly fulfil the evaluation's purpose and aim. Let us take the example further: in a given province, when the process ends, someone with access to all the data «leaks to the media» a ranking of centres placed in order of the mean mark that students obtained in each subject. The use of ranking and the information being publicised were not foreseen. The fact that this information was processed and diffused in this way clearly goes against the evaluation's consequential validity. It is no longer useful, and it loses its conclusion validity.

We believe that these generic and specific facets make up a reference concep-

tual framework that allows the validation analysis to be guided in any teaching evaluation plan, from its initial design to the assessment of its final validity.

## 4. Strategies to achieve validity

In addition a theoretical reflection on the concept of validity in evaluation, it is necessary to note the practical elements that make it possible to ensure that the evaluation is indeed valid. These practical elements are organised according to two types of strategies: those linked to planning the evaluation and those related to information management.

### 4.1. Planning the evaluation

Planning the evaluation is fundamental because it is the cornerstone of its validity (Jornet Meliá, Suárez Rodríguez, & Perales Montolío, 2003), and all the evaluation's design and management components must be contemplated to do this in a suitably way. These components are:

- Firstly, defining the evaluation's object and purpose. This involves defining all the elements that might affect the evaluated object to guarantee that the evaluation adequately does what is intended. To evaluate teachers, it is necessary to take a clear quality profile benchmark into account, and this is not generalisable, but is very dependent on the level being considered (early childhood, primary, secondary, post-compulsory, university), as well as the type of work

that teachers do (face-to-face or distance/virtual), because the quality profile is heavily dependent on teaching tasks. Defining the unit of analysis is the initial key element of validation because it is a matter of defining the tasks, competences, skills and attitudes that teachers must possess in order to be considered of quality. Thus, for instance, determining the differential weight of the emotional competence in the evaluation of early childhood teachers, or of university teachers' tutoring competences, and what the most relevant source of information is in each case (families or students themselves) is fundamental for defining the unit of analysis. Therefore, a complete, thorough, and coherent definition of the evaluation model that includes all of the dimensions involved recognised by the theoretical framework while respecting the cultural context in which it is used, helps to guarantee the evaluation's internal validity and, consequently, its credibility.

- Secondly, it is necessary to anticipate factors that facilitate and hinder the evaluation process. Taking care of the mediation process<sup>3</sup> in the evaluation is most important given the priority purpose of promoting acceptance of the evaluation, which must also be clear when establishing limits for using the evaluation information, and when adapting this to the purpose of the teaching evaluation in question. In

any case, the evaluation will always be an act of mediation between all of the groups involved (teachers, students, families, authorities and/or education consultants-administrators, society, etc.) in accordance with the level of education involved, and the end use for which the evaluation outcomes are intended. This is particularly important in teaching evaluations with summative consequences (recruitment, retention, promotion). Accordingly, the basic components to be considered when specifying the evaluation are to determine what is to be evaluated (in specific terms) and the purposes of the evaluation. Therefore, in the 1980s the idea of defining an *evaluation contract* was considered, which would have allowed the purpose of the evaluation to be explicitly determined by means of consensus, along with how the various groups involved would participate, the type of data to be collected, the analyses used on these data, and how outcomes would be used by specifying how each data type would be disseminated, whether or not each type of data would be kept confidential, and who would receive reports.

- Thirdly, and building on the previous point, caring for the participation, roles, and responsibilities of the involved groups is fundamental. We can ensure a quality evaluation if we can rely on the key in-

formants for each data type and if we are clear about the various roles that each involved group must play in the evaluation. Including all the relevant participants in planning and designing the evaluation gives the process credibility, which is the basic cornerstone of its utility (as previously pointed out: an evaluation will not be useful if it is not credible). Moreover, an evaluation is credible and useful when it has consequences and these match those planned. Even though the evaluation provides information, if its consequences are unclear for the person or group being evaluated or affected by the evaluation, it will be neither credible nor useful. To determine these consequences, it is vital to define well the participants and what their roles will be in the evaluation process.

- Finally, devising a work plan to carry out the evaluation itself is fundamental, and the work plan is valid when two conditions are ensured: the application's accuracy and its generalisability.
- 1. *The application's accuracy.* This is a case of ensuring that design of the work plan includes all of the elements (all of the competences) that we have defined as demonstrating teaching quality, thus ensuring both internal validity and construct validity. Similarly, all of the factors that condition or moderate the possibilities for teaching

must be included: input variables and the mediated (social) and immediate (institutional) contexts, and the ways by which they are carried out (personal processes). This is specified in the different evaluation phases:

- a) Data collection. Determining both how and when data will be collected without introducing disruptions or interferences that alter the reality and routine nature in which the educational situation being evaluated takes place. If, for example, we wish to evaluate whether a teacher delivers classes well, recording all classes and randomly selecting some would be technically ideal. However, from the ethical perspective, this presents problems that teachers often raise (data protection, academic freedom, protecting the teacher's image, etc.), which must doubtlessly be dealt with from a professional ethics perspective before opting for this type of solution. In any case, it is a matter of debate because we believe that the rights of the students being taught take priority over the rights of the people teaching them. So, it is necessary to deal with decisions of this kind using mediation processes.
- b) Data analysis: choosing suitable analyses to meet the eval-

uation's object and purpose, and analyses that allow statements to be made. All this has to be reflected in the evaluation plan. For example, to verify students' possible increased knowledge, a pretest-posttest design will be necessary, or a portfolio will be required to understand how teachers employ continuous evaluation systems. Apart from collecting portfolios, it will be necessary to determine a system that records narrations about how they are employed and how these uses are perceived.

- c) Interpreting data through contextual integration: apart from defining the teaching quality profile for each level of education and each educational modality, the evaluation must also be sensitive to the characteristics of the context (rural or urban setting, the socio-economic level of students' families, etc.). This is most important for modulating the interpretation of the results. However, this integration of contextual elements is very difficult because modulating the interpretation of results can only be based on a qualitative approach, on the evaluator's observation, which is one of the hardest evaluations to make.

- d) Use of data: having collected and analysed data, it is necessary to adapt their use to the purposes that the evaluation intends to meet. If formative evaluation is designed for improvement, it requires more detailed data analysis that identifies any gaps or shortcomings and the elements that can be improved.

2. *The translation's generalisability.*

The quality of the work plan as a basis of the evaluation's validity is reliant on this plan reflecting all the aspects involved in the concept of teaching quality, and the purpose and intention that inspired the design of the evaluation plan to make it a tool to achieve certain outcomes. Partial evaluations, such as those evaluating only mastery of the subject, planning competences, or ICT use, are inappropriate; it is only when all the components are coherently included that we can talk about a valid evaluation that fulfils the generality criterion of the translation of the evaluation's object, which is linked to construct validity.

Two evaluation situations are presented to as examples from higher education in Spain, which would probably not have occurred if this advice had been taken, especially if an evaluation contract existed. In Spain, evaluations of university teachers commenced in 1989 with research and teaching evaluations.

Research evaluation was designed in 1989 at the national level, with a formative purpose and was linked to pay rises (Tejedor & Jornet, 2008). The proposal was for them to be done on a 6-year cycle with voluntary participation by evaluating the five works that academics considered to be their best ones. As such, it was quite well accepted by this group of professionals. With time, however, this system has had summative consequences that go beyond the initially agreed salary supplements. Nowadays, many other consequences have derived from this evaluation — e.g. participating in thesis tribunals or teacher promotions, participating in supervising theses, consequences for teaching loads — that the initial design did not foresee.

Teaching evaluation in Spanish universities started with some isolated projects that were: a) experimental in nature and addressed to the design of instruments to adapt the evaluation's reality to the context of our universities (Aparicio, Tejedor, & Sanmartín, 1984); b) designed with institutional uses and purposes; e.g., in the 1986-87 academic year, the Universitat de València was closely linked to processes that improved teaching whose design was based on evaluation outcomes. This orientation helped to overcome teachers' considerable initial reluctance to be evaluated (and to be evaluated with students' opinions) and, thus, allowed the model to be implemented. Nevertheless, this formative task was lost with a change in the process's coordination, with training plans adapted to the evaluation's outcomes no longer being designed, and the assess-

ment was limited to summative use (consequences for pay rises), which has since intensified as it is also linked to promotion processes.

Such changes in how evaluation outcomes are used (by evaluators and lawmakers) are not ethical. It would have been impossible to change consequences if a professional approach had been taken or if a contract had specified the evaluation's uses and consequences. If those in charge had considered it necessary to introduce changes, a new agreement would have been needed. This is why real participation by the groups being evaluated is so important to ensure that the evaluation is accepted, that this acceptance is in writing, and that the exact conditions are explained, as happens in some countries. These ethical components are fundamental in evaluation processes with direct or indirect consequences for people's professional development. In short, the aforementioned consequential validity was evidently not fulfilled and, nowadays, such processes have become a meaningless administrative task for teachers and students alike, but which give the impression of an institution committed to using evaluation to monitor its actions.

#### 4.2. Strategies to ensure the evaluation's validity through data management

The second set of strategies that allow validity to be taken care of in teaching evaluation processes relate to data management, and these processes are structured around the successive evaluation phases.

1. Before the evaluation starts. The evaluation's planning must focus on thoroughly defining the evaluation's object; that is, the teacher profile we wish to evaluate. The detailed definition of all of its elements can guarantee that we will evaluate the key components for the decision-making process
2. While the evaluation process is underway, it is necessary to ensure that flexible planning is achieved. Prior planning must include a protocol, but it must also anticipate the elements to adapt to the situation using the factors that facilitate and hinder identified in advance (Nussbaum, 2006; Walker, 2003). Moreover, the data collection process must take care of the quality of the measuring instruments as a disciplinary field according to educational measuring principles. Despite evaluation's historical development, measurement and evaluation are still taken as synonyms in some contexts, even though they are not. Measurement is merely an instrumental process within the evaluation process, but it is a fundamental one, that must ensure the quality of both standardised instruments and evaluative techniques, which are qualitative but are also fundamental in teaching evaluation processes.
3. After completing the evaluation process, caring for the diffusion

and use of the evaluation is absolutely necessary, although these processes are normally neglected. For an evaluation to be credible and useful, it must be transmitted comprehensibly by explaining it well, along with its scope and limitations.

Finally, we stress that if the evaluation adopts these strategies to ensure validity throughout the process, and given the role that all the involved groups need to play according to its object and purpose, the evaluation would clearly play a mediating role in processes. This interpretation of evaluation as mediation can be made at several levels:

- Firstly, in the agreements reached before the evaluation. Following Cronbach's *negotiation* concept (1963), and according to what previous strategies indicate, the first evaluation phase involves negotiating with the involved groups to explicitly establish all of the elements included in the evaluation, including how the results are used and diffused.
- Planning the evaluation involves a new mediation process as a *compromise between what is desirable and what is possible*, between defining the consolidated object of evaluation within the theoretical framework and possibly transferring it to the evaluated natural situation, with the evaluation conditions being realistic, avoiding disruptions, and, as indicated in the strategies, making the

work plan flexible so it can adapt to the evaluated situation's real conditions. This also necessarily involves modulating the data interpretation and how data are used, which is conditioned by not only what the model foresees, but also what the evaluation process specifies.

- Specifically, in relation to how the evaluation is used and its consequences, as previously mentioned, the evaluation must always be an act of *mediation between reality and its interpretation* using the evaluation's outcomes.

## 5. Conclusions

In this work, we attempt to provide a global framework to conceptualise the validation of teaching evaluation processes. Work on validating teaching evaluation is still a very new area. Previous contributions have been made, but only in partial elements like validating instruments for teaching evaluation by students. It is now important to make efforts to provide an analysis that includes the validation of teaching in both the general validation elements and any specific elements that need to be established. Consequently, these specific elements will be established for evaluations according to the educational level and how teaching is delivered. In all cases, the teaching quality profile must be suitably specified, along with the expected teaching competences, and specific strategies must be developed in parallel to support teams that design and implement teaching evaluations to ensure the

quality of the evaluations. This is absolutely necessary because they have specific and sometimes serious consequences for the professional development of people, groups, and institutions.

Nonetheless, as we point out with this proposal, reflecting on validation processes in teaching evaluations involves starting a debate about:

- a) The status of evaluation in relation to research.
- b) Identifying an epistemological position to guide evaluation processes (considerations relating to validation processes will also undoubtedly derive from their positioning).
- c) Specific applications of validation that should be emphasised in accordance with how they are used in various teaching types (face-to-face vs. distance, levels of education, from early childhood to higher education).
- d) Evaluation typologies according to the evaluation's purpose (accountability vs. improvement-guided), etc.

It is, therefore, an initial proposal that must be developed and specified, and also debated, in relation to which approaches and methodological solutions might be most suitable in each case.

The reflection exercise we share here is based on our conviction that validating evaluation means making it useful to those involved, because if the evalua-

tion is not useful for anyone, it is basically not useful for anything. Validity is key to ensure evaluation processes make sense and it is the basis for ensuring justice and equity in the evaluation of any phenomenon in general, and teaching in particular. Unfortunately, all too often, we see evaluations with serious validity problems (from generic to specific facets) which are still proudly used by administrators and politicians who then use them to claim that education now has culture of evaluation. As we see it, working to develop a true evaluation culture is done by questioning the validity of evaluations, by analysing in-depth their designs and application, uses and consequences. Therefore, their basis is building a reference to help us to conceptualise their validity. We see this exercise as the basis of the ethical-deontological commitment that must dominate any evaluation process. To this end, we encourage researchers in education, especially specialists in educational measurements and evaluation, to join this commitment to develop a conceptual corpus to guide teaching evaluation quality.

## Notes

<sup>1</sup> The intention is to establish a cognitive parallelism with what takes place in measuring, in building standard measurement instruments where reliability is a necessary, but not a sufficient, condition to ensure validity.

<sup>2</sup> Many administrative processes used by Public Administrations employ models that are considered fair and equitable precisely because they are generic. However, this generic nature means that they often lose validity. This loss also takes place differentially among distinct centres because the evaluation model does not reflect the singularities of the different institutions.

<sup>3</sup> Traditionally, negotiation is regarded as a fundamental stage to adapt the evaluation plan. As we see it, this is a mediation process rather than a negotiation process. The aim is not that no party loses to carry out a process, but that all of the parties or groups involved in the evaluation understand that for the evaluation to provide elements for improvement, it must be regarded as a process that favours everyone.

## References

- Anguera, M. T., Blanco, A., Losada, J. L., & Portell, M. (2018). Pautas para elaborar trabajos que utilizan la metodología observacional. *Anuario de psicología*, 48 (1), 9-17.
- Aparicio, J. J., Tejedor, F. J., & Sanmartín, J. (1984). Reseñas de investigación. La enseñanza universitaria vista por los alumnos: un estudio para la evaluación de los cursos en la enseñanza superior. *Revista de Educación*, 273, 238-239.
- Bericat Alastuey, E. (1998). *La integración de los métodos cuantitativo y cualitativo en la investigación social: significado y medida*. Madrid: Ariel.
- Cronbach, L. J. (1963). Course improvement through evaluation. *Teachers College Record*, 64 (8), 672-683.
- Delgado Álvarez, C. (2014). *Viajando a Ítaca por mares cuantitativos: manual de ruta para investigar en grado y postgrado*. Salamanca: Amarú.
- García Mestanza, J. (2010). Propuesta de evaluación de la actividad docente universitaria en entornos virtuales de aprendizaje. *revista española de pedagogía*, 68 (246), 261-280.
- González-Sanmamed, M., Muñoz-Carril, P. C., & Santos-Caamaño, F. J. (2019). Key components of learning ecologies: A Delphi assessment. *British Journal of Educational Technology*, 50 (4), 1639-1655.
- González-Sanmamed, M., Sangrá, A., Souto-Seijo, A., & Estévez, I. (2018). Ecologías de aprendizaje en la era digital: Desafíos para la educación superior. *Publicaciones*, 48 (1), 11-38.
- González-Such, J. (1988). *Evaluación de la docencia universitaria*. Valencia: CSV.
- Guerra García, M., Leyva Barajas, Y. E., & Conzuelo Serrato, S. (2019). Factores contextuales que afectan la práctica docente desde la perspectiva de docentes en México. *Publicaciones*, 49, (1), 137-149. doi:10.30827/publicaciones.v49i1.9858



- Jornet Meliá, J. M. (2014, September). Facetas del concepto de validez en la evaluación de la Docencia. In *V Coloquio de la Red Iberoamericana de Investigadores sobre Evaluación de la Docencia (RIIED)*. Symposium conducted at the meeting of the Instituto de Investigación y Desarrollo Educativo of the Universidad Autónoma de Baja California (UABC), Ensenada, México.
- Jornet Meliá J. M., Bakieva, M., & Sánchez-Delgado, P. (in press). La cohesión social como objetivo de la educación: ¿Podemos especificar un modelo de calidad para realizar la evaluación de sistemas educativos? *Fronteiras: Journal of Social, Technological and Environmental Science*.
- Jornet Meliá, J. M., González-Such, J., Perales Montolío, M. J., Sánchez-Delgado, P., Bisquert, M., Bakieva, M., ... Ortega Gaite, S. (2017). Aproximaciones cualitativas para la definición y validación de constructos de instrumentos estandarizados de medida. In A. P. Costa, M. C. Sánchez-Gómez, & V. Martín Cillero (Eds.), *La Práctica de la investigación cualitativa: ejemplificación de estudios*. Salamanca: CIAIQ.
- Jornet Meliá, J. M., Suárez Rodríguez, J. M., & Perales Montolío, J. M., (2003). *Metodología de evaluación de programas de formación ocupacional y continua*. Valencia: Adeit.
- Jornet Meliá, J. M., Suárez Rodríguez, J. M., & Pérez Carbonell, A. (2000). La validez en evaluación de programas. *Revista de Investigación Educativa*, 18 (2), 341-356.
- Lorenzo Quiles, O. (2011). Análisis cualitativo de textos sobre multi e interculturalidad. *DEDiCA. Revista de Educação e Humanidades*, 1, 535-546.
- Manzi, J., González, R., & Sun, Y. (Eds.) (2011). *La evaluación docente en Chile*. Santiago de Chile: MIDE-UC.
- Martínez-Rizo, F. (Coord.) (2015). *Las pruebas ENLACE y EXCALE. Un estudio de validación. Cuadernos de Investigación*, 40. México, DF: Instituto Nacional para la Evaluación de la Educación.
- Martín-García, A. V., Sánchez-Gómez, M. C., & Costa, P. (2019). Percepción de Blended Learning en profesores universitarios de distintos ámbitos disciplinares. *Revista Lusófona de Educação*, 44 (44), 117-133.
- Mertens, D. M. (2015). Mixed methods and wicked problems. *Journal of Mixed Methods Research*, 9 (1), 3-6.
- Nava, M., & Rueda, M. (2014). La evaluación docente en la agenda pública. *Revista Electrónica de Investigación Educativa*, 16 (1), 1-11.
- Nussbaum, M. C. (2006). Education and democratic citizenship: capabilities and quality education. *Journal of Human Development*, 7 (3), 385-395.
- Ortega Gaite, S., Tejedor Mardomingo, M., & Perales Montolío, M. J. (2019, September). La distribución económica y desigualdad en los manuales escolares de educación para la ciudadanía y derechos humanos en clave de educación para el desarrollo. In *Educación Crítica: hacia una práctica inclusiva y comprometida socialmente (EDUCRÍTICA 2019)*. Symposium conducted at the international meeting of the Faculty of Education of the Universidad de León (ULE), León, Spain.
- Rueda-Beltrán, M. (Coord.) (2019). *Las condiciones institucionales de la docencia en universidades iberoamericanas*. Temuco: Red Iberoamericana de Investigadores sobre Evaluación de la Docencia (RIIED), Universidad de La Frontera.
- Ruiz-Primo, M. A., Jornet Meliá, J. M., & Backhoff Escudero, E. (2006). *Acerca de la Validez de los Exámenes de la Calidad y el Logro Educativos (Excale)*. Cuadernos de Investigación (México), 20. Ciudad de México: Instituto Nacional para la Evaluación de la Educación. Retrieved from <https://www.inee.edu.mx/publicaciones/acerca-de-la-validez-de-los-examenes-de-la-calidad-y-el-logro-educativas-excale-no-20/> (Consulted on 2019-08-25).
- Rutman, L. (1984). *Evaluation research methods: a basic guide*. California: Sage Publications.
- Sánchez-Gómez, M. C., Martín-Cilleros, M. V., Costa, A. P., & Peñalvo, F. J. G. (2018). Posicionamiento de la investigación en Ciencias Sociales. *RISTI - Revista Ibérica de Sistemas e Tecnologías de Informação*, 28, 102-113.
- Sánchez-Gómez, M. C., Rodrigues, A. I., & Costa, A. P. (2018). Desde los métodos cualitativos hacia los modelos mixtos: tendencia actual de investigación en ciencias sociales. *RISTI - Revista Ibérica de Sistemas e Tecnologías de Informação*, 28.
- Sandín Esteban, M. P. (2000). Criterios de validez en la investigación cualitativa. De la objetividad a la solidaridad. *Revista de Investigación Educativa*, 18 (1), 223-242.
- Solano-Flores, G., Contreras-Niño, L. A., & Backhoff, E. (2006). Traducción y adaptación de pruebas: Lecciones aprendidas y recomendacio-

- nes para países participantes en TIMSS, PISA y otras comparaciones internacionales. *Revista Electrónica de Investigación Educativa*, 8 (2).
- Solano-Flores, G., & Milbourn, T. (2016). Capacidad evaluativa, validez cultural y validez consecuen- cial en PISA. *RELIEVE Revista Electrónica de Investigación y Evaluación Educativa*, 22 (1).
- Taut, S., & Palacios, D. (2016). Interpretaciones no intencionadas e intencionadas y usos de los resultados de PISA: Una perspectiva de validez consecuen- cial. *RELIEVE Revista Electrónica de Investigación y Evaluación Educativa*, 22 (1).
- Tejedor, F. J. (2004). Investigación educativa: ¿Ha- cia dónde vamos? In L. Buendía, D. González, & T. Pozo (Coords.), *Temas fundamentales en la investigación educativa* (pp. 63-107). Ma- drid: La Muralla.
- Tejedor, F. J., & García Valcárcel, A. (2010). Evalua- ción del desempeño docente. *revista españo- la de pedagogía*, 68 (247), 439-459.
- Tejedor, F. J., & Jornet Meliá, J. M. (2008). La eva- luación del profesorado universitario en Espa- ña. *REDIE Revista Electrónica de Investiga- ción Educativa*, 10, special issue 2008, 1-29.
- Vedung, E. (1995). Utilización de la evaluación de programas. In R. Fernández Ballesteros (Ed.), *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud* (pp. 321-338). Madrid: Síntesis.
- Walker, M. C. (2003). Framing social justice in edu- cation: what does the «capabilities» approach offer? *British Journal of Educational Studies*, 51, 168-187.
- Weiss, C. (1998). *Evaluación*. Nueva Jersey: Pren- tice-Hall.

## Authors' biographies

**Jesús Miguel Jornet Meliá** is a PhD in Philosophy and Education Sciences / Pedagogy from the Universitat de València and a Professor of Educational Measure- ment and Evaluation. He was the Director of the Research Methods and Diagnosis in Education Department (1993-1999) at the same university. Currently he coordinates

the GemEduCo group: [www.uv.es/gem](http://www.uv.es/gem). Main research interests: design of meas- urement and evaluation tools, evaluation of competencies, programmes, teachers and education systems.

 <https://orcid.org/0000-0001-6905-497X>

**María Jesús Perales Montolío** has a PhD in Philosophy and Educational Sciences from the Universitat de València. She is currently Associate Professor in the Department of Research Methods and Diagnosis in Education; Director of the Official Master in Social and Ed- ucational Action, and Member of the GEM-Educo Research Group. Main re- search intersts: design of measurement and evaluation tools, evaluation of com- petencies, programmes, teachers and ed- ucation systems.

 <https://orcid.org/0000-0003-2033-2750>

**José González Such** is a PhD in Philosophy and Educational Sciences. He is currently Professor in the area of Research Methods and Diagnosis in Educa- tion and Director of the Master in Special Education at the University of Valencia, Member of the GemEduco group and Coord- inator of the Educational Innovation and Teaching Quality Network InnovaMIDE. Lines of research: design of measurement and evaluation instruments, evaluation of competencies, programs, teachers and ed- ucational systems; educational innovation.

 <https://orcid.org/0000-0001-9086-6446>

# Table of Contents

## Sumario

### Editorial

---

**José Antonio Ibáñez-Martín**

An educational consideration on the pandemic:  
endure...and progress

*Una consideración educativa sobre la pandemia: resistir...  
y adelantar*

181

### Studies

---

**Bernardo Gargallo López, Fran J. García-García,  
Inmaculada López-Francés, Miguel Ángel Jiménez  
Rodríguez, & Salomé Moreno Navarro**

The learning to learn competence: An assessment  
of a theoretical model

*La competencia aprender a aprender: valoración de  
un modelo teórico*

187

**Miguel A. Santos Rego, María José Ferraces Otero,  
Ígor Mella Núñez, & Ana Vázquez-Rodríguez**

University, civic-social competences, and the  
labour market

*Universidad, competencias cívico-sociales y mercado de  
trabajo*

213

**Jesús Miguel Jornet Meliá, María Jesús Perales  
Montolio, & José González-Such**

The concept of validity of teaching evaluation  
processes

*El concepto de validez de los procesos de evaluación  
de la docencia*

233

**Fernando Acevedo Calamet**

Explanatory factors for dropout from higher education  
in unfavourable socio-academic contexts

*Factores explicativos del abandono de los estudios en la educación  
superior en contextos socio-académicos desfavorables*

253

### Notes

---

**María-Carmen Ricoy, & Cristina Sánchez-Martínez**

A systematic review of tablet use in primary education

*Revisión sistemática sobre el uso de la tableta en la etapa de  
educación primaria*

273

**Jaime Carcamo-Oyarzun, & Christian Herrmann**

Construct validity of the MOBAK test battery for the  
assessment of basic motor competencies in primary  
school children

*Validez de constructo de la batería MOBAK para la evaluación  
de las competencias motrices básicas en escolares de  
educación primaria*

291

**Concha Iriarte Redín, Sara Ibarrola-García, & Maite Aznárez-Sanado**

Proposal for a school mediation evaluation tool (MEQ)

*Propuesta de un instrumento de evaluación de la mediación escolar (CEM)*

**Bruno Echaury Galván, & Silvia García Hernández**

Translating in colours: Translation as an assessment tool for reading comprehension in English as a second language

*Traducir en colores: la traducción como herramienta de evaluación de la comprensión lectora en inglés como lengua extranjera*

## Book reviews

---

- 309 **Esteban-Bara, F. (2019).** *La universidad light: Un análisis de nuestra formación universitaria [The light university: An analysis of our university education]* (José L. González-Geraldo). **Ruiz-Corbella, M., & García-Gutiérrez, J. (Eds.) (2019).** *Aprendizaje-Servicio. Los retos de la evaluación [Service learning: The challenges of evaluation]* (Andrea Muñoz Villanueva). **Lafforgue, L. (2019).** *Recuperemos la escuela [Reclaiming the school]* (Beatriz Gálvez). **Ferraces-Otero, M. J., Godás-Otero, A., & García-Álvarez, J. (2019).** *Cómo realizar un estudio científico en ciencias sociales, de la educación y de la salud [How to do a scientific study in the social, educational, and health sciences]* (Carolina Rodríguez-Llorente). 347

This is the English version of the research articles and book reviews published originally in the Spanish printed version of issue 276 of the **revista española de pedagogía**. The full Spanish version of this issue can also be found on the journal's website <http://revistadepedagogia.org>.



ISSN: 0034-9461 (Print), 2174-0909 (Online)

<https://revistadepedagogia.org/>

Depósito legal: M. 6.020 - 1958

INDUSTRIA GRÁFICA ANZOS, S.L. Fuenlabrada - Madrid