



Validez y confiabilidad en la evaluación del aprendizaje mediante las metodologías activas

Validity and reliability in student learning evaluation throughout active methodologies

ib **Dra. María del R. Medina-Díaz** es docente e investigadora de la Universidad de Puerto Rico (Puerto Rico) (maria.medina2@upr.edu) (<https://orcid.org/0000-0002-0197-2480>)

ib **Dra. Ada L. Verdejo-Carrión** es docente e investigadora de la Universidad de Puerto Rico (Puerto Rico) (ada.verdejo@upr.edu) (<https://orcid.org/0000-0001-9522-3116>)

Recibido: 2020-02-01 / **Revisado:** 2020-06-01 / **Aceptado:** 2020-06-09 / **Publicado:** 2020-07-01

Resumen

Las exigencias en la educación universitaria demandan cambios en las estrategias de enseñanza y en las técnicas y los instrumentos que contribuyen a evaluar el aprendizaje estudiantil. Las metodologías activas, como parte de estas estrategias, facilitan el desarrollo de determinados aprendizajes o competencias, mediante situaciones o problemas vinculados con el mundo laboral y social. Esto requiere replantear, planificar y orientar la enseñanza centrada en el estudiantado y utilizar técnicas e instrumentos para recoger información que conduzcan a emitir juicios apropiados, certeros y justos de los aprendizajes. Sin embargo, no se ha prestado mucha atención a la validez y la confiabilidad de las interpretaciones de las puntuaciones o la información recopilada con estos instrumentos para el uso propuesto, según se desprende de las publicaciones revisadas. El propósito de este trabajo es aportar a la discusión acerca de la validez y confiabilidad de las puntuaciones o la información recopilada con los instrumentos aplicados en las aulas universitarias. Se consultaron varias publicaciones especializadas y se presentan algunas recomendaciones acerca de las fuentes de evidencia para sustentar la validez y la confiabilidad. Como mínimo, se sugiere la evidencia relacionada con el contenido y la consistencia de las puntuaciones o la información, al emitir juicios y tomar decisiones que afectan al estudiantado. Se concluye que se necesita mayor prudencia en las interpretaciones e inferencias de los aprendizajes, si no existe suficiente evidencia de la validez.

Descriptores: Metodologías activas, evaluación de estudiantes, validez, confiabilidad, evaluación del aprendizaje, técnicas de evaluación.

Abstract

The demands on university education call for changes in teaching strategies and on the evaluation of student learning. Active methodologies are part of these strategies, which facilitate the development of student learning or competencies, through situations or problems close to the real world and to a professional career. These require to rethink and plan teaching as student-centered, as well as to use techniques and instruments for collecting valid and reliable information, that leads to an appropriate and a fair evaluation of student learning. However, the evidence of validity and reliability of the interpretations of instrument scores or information has not had enough attention, according to the literature reviewed. The purpose of this paper is to discuss the validity of the interpretation and reliability of the scores or the information collected through classroom assessment tools in universities. According to some publications, a set of recommendations is provided with sources of evidence that underpin validity and reliability. At minimum, evidence related to content validity and to internal consistency of the scores or information collected should be taken into account, when making judgments and decisions that affect the students. It is concluded that greater prudence is needed in the interpretations and inferences of learning, if there is insufficient validity evidence.

Keywords: Active methodologies, student evaluation, validity, reliability, student learning, evaluation techniques.

1. Introducción

En pasadas tres décadas, las exigencias del público, los gobiernos y las agencias acreditadoras en cuanto a la educación universitaria demandan del estudiantado aprendizajes variados y complejos para el desempeño en el mundo laboral y a lo largo de la vida (Erwin, 1991; Huba & Freed, 2000; Krzykowski & Kinser, 2014; McClarty & Gaertner, 2015; Pozo & Pérez-Echeverría, 2009; UNESCO, 1998). Estas expectativas suponen ciertos cambios en las estrategias de enseñanza y en las técnicas y los instrumentos que recopilan información para evaluar el aprendizaje estudiantil. Implica mirar de manera distinta el aprendizaje para atender a las características, los procesos y estilos de aprender del estudiantado y proveer actividades instruccionales, donde construyan conocimientos y destrezas, partiendo de los previos, y oportunidades de involucrarse activamente, demostrando lo que han aprendido y así evaluar su desempeño (Brookhart, 2004; Erwin, 1991; Hortigüela-Alcalá et al., 2015; Huba & Freed, 2000; López-Pastor & Sicilia-Camacho, 2016). Con esta perspectiva, el énfasis radica en cómo aprender o elaborar estructuras mentales y procesos de pensar y actuar a fin de desarrollar y lograr los aprendizajes esperados. Estos son múltiples y se integran en las dimensiones cognitivas, afectivas, psicomotoras y sociales del desarrollo académico y personal del estudiantado en los distintos contextos educativos. Comúnmente, estos aprendizajes se enuncian como objetivos de enseñanza (*“learning targets”*, Stiggins, 2017, p.11); resultados o metas del aprendizaje (*“learning outcomes”*, *“learning goals”*, Huba & Freed, 2000, p. 94, p. 9, respectivamente) y competencias (Baartman et al., 2006; De la Orden, 2011; Epstein, 2007; Fernández March, 2006; García-Merino et al., 2016; Goñi Zabala, 2005; Olmos-Miguelañez & Rodríguez-Conde, 2010; Pozo & Pérez-Echeverría, 2009; Voorhees, 2001).

Las metodologías activas o auténticas, como estrategias de enseñanza, sirven de vehículo para facilitar su desarrollo y logro, mediante

situaciones o problemas similares a los que se confrontan en el campo profesional y la sociedad. Su aplicación, requiere replantear, planificar y orientar la enseñanza de modos distintos, ubicando al o la estudiante como centro, y alineando las técnicas y los instrumentos para la evaluación. Una técnica de (o para la) evaluación se refiere al conjunto de procedimientos o acciones planificadas para recoger información acerca del aprendizaje; mientras que un instrumento es el objeto o medio específico con el cual ésta se aplica. Aquí la información abarca puntuaciones, selecciones, anotaciones, comentarios u otras maneras que muestren respuestas u observaciones. Medina Díaz y Verdejo Carrión (2019) las clasifican en cuatro grupos, con los instrumentos asociados: (a) pruebas (e.g., pruebas objetivas y subjetivas); (b) observación (e.g., listas de cotejo, escalas de categorías y rúbricas); (c) comunicación personal (e.g., entrevista, cuaderno) y (d) tareas de ejecución (e.g., proyecto, portafolio). Angelo y Cross (1993), Barkley y Major (2016), Suskie (2009) y Weimer (2013) presentan múltiples ejemplos de estas técnicas para la evaluación del aprendizaje atemperadas al contexto universitario. Se confía en que el profesorado las conoce y utiliza aquellas que armonizan con los objetivos de un curso, las estrategias de enseñanza que emplea y la diversidad de características del estudiantado, para producir información apropiada de sus aprendizajes (Banta et al., 1996; Black & William, 1998a, 1998b; Bennett, 2011; Davies & Taras, 2018; López-Pastor & Sicilia-Camacho, 2016; Newble & Cannon, 1991; Olmos-Miguelañez & Rodríguez-Conde, 2010; Rawlusk, 2018; Webber, 2012). Conviene aclarar que una técnica o un instrumento no son exclusivos de un tipo de evaluación (e.g., diagnóstica, formativa o sumativa), sino que es el profesor o la profesora quién determina el propósito y uso de las puntuaciones o de los resultados. En esto, precisamente, descansa la validez de las interpretaciones de las puntuaciones o la información para el uso designado. Esto, con la expectativa de que mejore sustancialmente la calidad de la experiencia académica y los aprendizajes o las



competencias del estudiantado universitario. La Tabla 1 presenta tres ejemplos de metodologías activas y posibles técnicas e instrumentos de recopilación de información para evaluar el logro de determinados aprendizajes.

Tabla 1. Ejemplos de metodologías activas, técnicas e instrumentos para recopilar información

Metodologías activas	Definición	Técnicas (e instrumentos)
Proyecto	Conjunto de actividades que el estudiantado realiza, de manera individual o grupal, durante un tiempo prolongado, con el propósito de atender un problema o asunto y produce un objeto, prototipo, informe oral o escrito.	Observación sistemática (Lista de cotejo) Comunicación personal (Bitácora de aprendizaje) Tarea de ejecución (Rúbrica)
Solución de problemas	Proceso mediante el cual el estudiantado realiza una serie de acciones y toma decisiones integrando conocimientos, destrezas y actitudes para responder o resolver un problema o una situación real y para la cual no hay una solución única	Comunicación personal (Hojas de reflexión y autoevaluación) Tarea de ejecución (Rúbrica)
Aprendizaje cooperativo	Grupo pequeño de estudiantes en el que todos interactúan y participan para un asunto, realizar una tarea o alcanzar una meta común.	Observación sistemática (Escala de categorías) Comunicación personal (Bitácora de aprendizaje, Hojas de autoevaluación y coevaluación)

La evaluación del aprendizaje conlleva emitir un juicio fundamentado, basado en información apropiada y pertinente acerca de los diversos aprendizajes desarrollados y logrados. Por lo tanto, uno de los grandes retos es recopilar y combinar información, tanto cuantitativa como cualitativa, recogida con múltiples instrumentos y en distintos momentos. Sin embargo, bajo el manto de la informalidad y la rapidez con que suele ocurrir la evaluación en las aulas, no se considera la validez de las interpretaciones de las puntuaciones o de la información recogida, así como la confiabilidad. Tal vez por esto, han tenido una mención exigua en publicaciones, en inglés, dedicadas a técnicas de *assessment* en la educación superior (Angelo & Cross, 1993; Barkey & Major, 2016; Huba & Freed, 2000; Suskie, 2009; Wolf et al., 2012), así como en las

investigaciones acerca de las prácticas de evaluación del profesorado universitario en varios países (Alquraan, 2012; Andreu-Andrés & Labrador-Piquer, 2011; Bearman et al., 2017; Brown & Atkins, 1988; Erwin, 1991; Gilles, Detroz & Blais, 2010; Goubeaud, 2010; Goubeaud & Yan, 2004; Hernández, 2012; Hortigüela Alcalá et al., 2017; Pereira & Flores, 2016; Yükseli & Gündüz, 2017). También, cabe la posibilidad que, en aras de la confianza y la libertad académica, la validez y la confiabilidad se dan por sentado, cuando el profesorado elabora o utiliza uno o más instrumentos para aplicarlos al estudiantado; sin contar con evidencia suficiente que lo sustente y el cuidado de ciertos elementos técnicos en su construcción y aplicación (Esteve Zarazaga, 2007; Gil Flores, 2005; Jacobs & Chase, 1992; O'Hagan, 2014; Poskanzer, 2002). La evaluación



del aprendizaje integrada con metodologías activas en la enseñanza invita a considerar la calidad de los instrumentos aplicados en el nivel universitario, en particular la validez y la confiabilidad de las puntuaciones o de la información que se obtiene. Este ensayo tiene como propósito principal aportar a la discusión de estos dos aspectos.

2. Metodología

Con este fin, se revisaron varias publicaciones que favorecen o debaten el cumplimiento con las consideraciones de la validez y la confiabilidad de las interpretaciones de la información recopilada con los instrumentos utilizados para evaluar el aprendizaje (American Educational Research Association et al., 2018; Brookhart, 2003, 2007; Cizek, 2009, 2015; Joint Committee on Standards for Educational Evaluation, 2018; Moss, 2003). Los *Estándares para pruebas educativas y psicológicas*, publicados por las organizaciones American Educational Research Association, American Psychological Association y National Council on Measurement in Education (2018) apuntan hacia las exigencias que deben cumplir los instrumentos que se construyen y administran a gran escala o con fines comerciales para asegurar la validez de las interpretaciones de las puntuaciones con el uso propuesto. Estos ofrecen un marco de referencia o guía para garantizar que se aborden cuestiones relevantes en la construcción de instrumentos educativos y psicológicos y brindan una base para revisarlos y criticarlos (American Educational Research Association et al., 2018, p.1). El Joint Committee on Standards for Educational Evaluation (2015), atiende el asunto, referente a lo que ocurre en las aulas escolares, en el estándar “*Q4 Reliability and Validity*”: “*Classroom assessment practices should provide consistent, dependable and appropriate information that supports interpretations and decisions about each student’s knowledge and skills*” (Localización 643).

Desde nuestra perspectiva, no se pretende la misma amplitud y rigor estadístico que en los procedimientos de validación empleados en

las pruebas de aprovechamiento estandarizadas y otros instrumentos aplicados en la práctica y la investigación educativa y psicológica. Sin embargo, los que se utilizan en las universidades, también, deben generar información cuantitativa y cualitativa apropiada acerca del aprendizaje estudiantil, conforme al propósito, el contenido, las estrategias de enseñanza y el tamaño de los grupos de estudiantes y otras circunstancias, así como a las inferencias y acciones derivadas. Una de estas inferencias podría ser que el o la estudiante haya logrado la competencia de solución de problemas, aplicando estadísticas descriptivas. Moss (2003), por su parte, plantea que la noción de validez no es suficiente para la práctica en las aulas y que se debe reconceptualizar. Aboga por enfoques interpretativos (e.g., basados en la teoría sociocultural y la hermenéutica) para manejar la información que se recopila continuamente. Brookhart (2003) propone el desarrollo de una teoría de medición basada en la sala de clases (“*classroometric*”, p.8). Ante estos planteamientos, destacamos los conceptos de validez y confiabilidad y las fuentes de evidencia que les sirven de apoyo. Además, presentamos varias recomendaciones acerca de la evidencia pertinente para interpretar las puntuaciones y otra información recopilada con los instrumentos.

3. Validez y confiabilidad

3.1. Validez

La validez se define al “grado en que la evidencia y la teoría respaldan las interpretaciones de los puntajes de una prueba o instrumento de medición para los usos propuestos” (American Educational Research Association et al., 2018, p.11). En otras palabras, abarca un juicio acerca de la interpretación de las puntuaciones o de la información obtenida con un instrumento, a la luz de la evidencia de distintas fuentes que le sirven de sostén. Estas se relacionan con el contenido, el proceso de respuesta, la estructura interna, con otras variables y con las consecuencias



(American Educational Research Association et al., 2018). Esta visión se fundamenta en un concepto unificador de la validez, contemplando que la evidencia integrada de estas fuentes contribuye a la validez relacionada con el constructo, de modo que sustente teórica y empíricamente que el instrumento lo mide o representa de manera apropiada y conduzca a inferencias y acciones adecuadas (Messick, 1989). Un instrumento que

se utilice en las aulas universitarias, requiere evidencia relevante que apoye las interpretaciones y los usos de las puntuaciones o la información obtenida. La Tabla 2 resume algunos de los procedimientos vinculados con las cinco fuentes de evidencia, de acuerdo con Cizek (2009), Medina-Díaz y Verdejo-Carrión (2019), McMillan (2008) y Nitko y Brookhart (2011).

Tabla 2. Recomendaciones para la validez y la confiabilidad de la interpretación de las puntuaciones o la información recogida con un instrumento para la evaluación

Fuentes de evidencia	Recomendaciones
Validez relacionada con el contenido	<ul style="list-style-type: none"> • Preparar las especificaciones del instrumento con los aprendizajes o los temas de contenido. • Proveer un número suficiente de ítems o tareas asociadas con los aprendizajes. • Construir los instrumentos, siguiendo las recomendaciones de fuentes de referencias reconocidas en el campo de la evaluación del aprendizaje. • Revisar la claridad de los ítems o las tareas del instrumento, así como las instrucciones. • Asegurar que el vocabulario, la estructura gramatical, el lenguaje y el formato de los ítems o las tareas sea adecuado para el estudiantado. • Determinar la correspondencia entre los ítems o las tareas del instrumento con los aprendizajes y los temas de contenido que intentan representar.
Validez relacionada con el proceso de respuesta	<ul style="list-style-type: none"> • Verificar la concordancia entre las respuestas ofrecidas a los ítems o las tareas y los aprendizajes (e.g., estrategias o procesos cognitivos). • Identificar los procesos cognitivos, las destrezas o estrategias necesarias para contestar los ítems o las tareas. • Entrevistar a un grupo de estudiantes, inmediatamente después de administrar un instrumento, para conocer las estrategias o procesos empleados en contestar. • Proveer tiempo y acomodos necesarios cuando se requiere procesos cognitivos o destrezas complejas. • Solicitar al estudiantado que explique el trabajo realizado o que muestren los pasos o procedimientos para llegar a la respuesta.
Validez relacionada con la estructura interna	<ul style="list-style-type: none"> • Analizar la consistencia de las respuestas a los ítems o las tareas de un instrumento asociadas al mismo aprendizaje o contenido. • Verificar las coincidencias entre las puntuaciones, así como las discrepancias con trabajos, previamente calificados con un instrumento.
Validez relacionada con otras variables	<ul style="list-style-type: none"> • Comparar los resultados de un instrumento, antes y después de la discusión de un tema o del proceso de enseñanza-aprendizaje. • Contrastar la ejecución en distintos instrumentos que representan los mismos y distintos aprendizajes. • Identificar las características, las experiencias y las necesidades educativas del estudiantado que responde o con quien se utiliza el instrumento.



Fuentes de evidencia	Recomendaciones
Consecuencias	<ul style="list-style-type: none"> • Identificar los efectos o las repercusiones del uso del instrumento y la información obtenida. • Revisar las interpretaciones y decisiones tomadas acerca del aprendizaje y la instrucción, de acuerdo con la información recopilada. • Asociar la interpretación de la información del instrumento con las decisiones correspondientes. • Solicitar reacciones o comentarios del estudiantado acerca de los instrumentos y las calificaciones.
Confiabilidad	
Consistencia en las respuestas o la ejecución	<ul style="list-style-type: none"> • Contar con suficiente información acerca de los aprendizajes desarrollados y logrados, mediante diversos instrumentos. • Disponer de una clave o guía para revisar las contestaciones a los ítems o la ejecución en las tareas. • Establecer e informar acerca de los criterios e indicadores que se utilizarán para observar o calificar la ejecución en las tareas. • Presentar ejemplos de respuestas esperadas y de trabajos en los distintos niveles de ejecución. • Revisar o calificar todas las contestaciones del grupo de estudiantes a una pregunta para elaborar la respuesta, antes de pasar a otra. • Proveer dos o más ocasiones para contestar preguntas o realizar tareas, relacionadas con los aprendizajes esperados y comparar la ejecución. • Utilizar dos o más personas (profesor, profesora y estudiante) para calificar la ejecución, comparar las puntuaciones y calcular el porcentaje de acuerdo. • Describir los procedimientos para corregir o calificar las respuestas o los trabajos.

La evidencia de la validez relacionada con el contenido comienza a acumularse desde que se identifica el propósito y los aprendizajes que se pretenden representar en un instrumento. Su construcción requiere delimitar los aprendizajes, los temas de contenido, la cantidad y el tipo de ítems como parte de las especificaciones que le sirven de base. Por ejemplo, cuando se trata de una prueba de aprovechamiento objetiva, se acostumbra colocar usando el formato de una tabla y se ubican los aprendizajes esperados o los temas de contenido en las filas o las columnas o simplemente, los objetivos con los ítems asociados (Medina-Díaz & Verdejo-Carrión, 2019). En este y otros instrumentos que se diseñen, como pruebas de ejecución, se espera una concordancia entre los ítems o las tareas, los objetivos, las estrategias de enseñanza, el énfasis y el tiempo dedicado a la discusión del contenido (i.e., validez instruccional). A partir de las especificaciones se

elaboran los ítems o las tareas, según el número y tipo adecuado, atendiendo a los aprendizajes, el tiempo y otras condiciones del estudiantado y el contexto universitario. Para elaborar las preguntas tipo ensayo o de discusión, las cuales se utilizan con mucha frecuencia en este contexto, también se considera la organización y relación entre los temas en la respuesta ideal (Brown, 2010; Medina-Díaz & Verdejo-Carrión, 2019).

Los ítems o las tareas, junto con las instrucciones, son las piezas principales de un instrumento. Por lo tanto, su selección y elaboración merecen esmero. Nitko y Brookhart (2011), Haladyna (1997), Mateo y Martínez (2008); Medina-Díaz y Verdejo-Carrión (2019) presentan numerosas recomendaciones para crearlos. Entre las generales se encuentran: (a) claridad en el vocabulario empleado, (b) sencillez en la estructura gramatical y (c) evitar incluir dos ideas y lenguaje sexista, ofensivo o discriminatorio.



rio. Asimismo, se redactan las respuestas correctas o esperadas a los ítems o tareas. Un asunto crucial es asegurarse que los ítems o las tareas requieran mostrar aprendizajes cognoscitivos, al menos de comprensión y aplicación de conceptos, acciones y procedimientos. Esto supone que el profesorado ha utilizado estrategias de enseñanza que han ayudado a desarrollarlos.

En la preparación de un instrumento también se toma en cuenta la secuencia, la edición y el despliegue físico. Esto se refleja en la apariencia y organización, así como en la ausencia de errores ortográficos en los ítems y las instrucciones. Lo anterior aplica tanto a un instrumento en formato impreso como electrónico. Estas acciones, sin duda, abonan a la evidencia relacionada con el contenido del instrumento, la cual tiene mayor envergadura en la evaluación del aprendizaje estudiantil.

Conviene recordar que la revisión de la representatividad y la relevancia de los ítems o las tareas de una prueba de aprovechamiento u otro instrumento estandarizado, dependen del juicio de personas conocedoras o familiarizadas con el contenido, los aprendizajes y el grupo estudiantil de interés (American Educational Research Association et al., 2018). En las aulas universitarias, esta labor la realiza solamente el profesor o la profesora, quien decide acerca de los aprendizajes que se van a representar, los temas del contenido y los ítems o las tareas del instrumento. Además, cabe la posibilidad de la sub- o sobre-representación de algún objetivo o tema del curso. Para minimizar esto, un o una colega podría ayudar a revisar la coherencia del instrumento con las especificaciones y la ambigüedad en las preguntas, antes de administrarlo.

La evidencia acerca del proceso de respuesta incluye establecer la afinidad entre los ítems o las tareas del instrumento con los aprendizajes cognoscitivos, afectivos y psicomotores o las destrezas que se requieran para contestarlos o completarlos. Para recopilar evidencia acerca de los distintos procesos cognitivos o destrezas de pensamiento u otras es necesario, al menos,

identificar cómo los ítems o las tareas las modelan y capturan la información relevante. Por ejemplo, para realizar una tarea de ejecución en Matemáticas se espera que el o la estudiante emplee una estrategia general o heurística (e.g., realizar un diagrama o dibujo) para contestarla. La profesora o el profesor podrían entrevistar a varios estudiantes y solicitarles que expliquen su razonamiento para verificar si contestaron aplicando dicha estrategia en la tarea o solución al problema en cuestión. También puede observar la ejecución de un grupo de estudiantes realizando cierta tarea para indagar si el procedimiento que llevaron a cabo es el contemplado.

La evidencia de la estructura interna conlleva mostrar la cohesión de los ítems o las tareas del instrumento en representar lo que se propone (e.g., aprendizajes o temas de contenido), y la consistencia en las respuestas. Nitko y Brookhart (2011) sugieren prestar atención a los patrones de respuesta a los ítems o las tareas y a la concordancia con los de otros instrumentos administrados. Si se trata de pruebas, se aconseja analizar y comparar las respuestas de varias preguntas o tareas del mismo contenido, para determinar la consistencia de los resultados (McMillan, 2008). La evidencia de la relación con otras variables concierne a la asociación entre las respuestas a los ítems y las puntuaciones del instrumento con variables externas (llamados criterios). Esta evidencia es pertinente cuando las puntuaciones en un instrumento sirven como indicador del desempeño en otras variables (e.g., promedio académico, puntuaciones en una prueba de razonamiento lógico). En la evaluación del aprendizaje esto se podría aprovechar, primero, identificando las características, las experiencias y la diversidad de necesidades educativas del estudiantado que responde o con quien se utiliza el instrumento y luego, buscar similitudes y diferencias en la ejecución del estudiantado en el instrumento en cuestión y en otros administrados (McMillan, 2008). También se podrían comparar las observaciones en las tareas del estudiantado, antes de iniciar la discusión del tema y al terminarla. Las



discrepancias sugieren posibles cambios en los aprendizajes desarrollados. Asimismo, sería útil verificar la ejecución en varias tareas dirigidas a demostrar aprendizajes similares o asociados.

La evidencia relacionada con las consecuencias considera los posibles efectos o las repercusiones que tienen los instrumentos y el uso de la información recopilada. Exige, por ejemplo, documentar cómo se interpretaron las puntuaciones (e.g., interpretación normativa o criterial), para qué se usaron y cuáles fueron las consecuencias (e.g., aumentar la motivación o el tiempo de estudio, reducir la cantidad de fracasos). Además, se podría indagar el impacto en el aprendizaje estudiantil de una tarea de ejecución dirigida a realizar un trabajo de investigación o práctico en una comunidad (Ricoy & Fernández-Rodríguez, 2013). Una entrevista con el o la estudiante o una reflexión escrita en el trabajo serviría de evidencia. McMillan (1997) y Taylor y Nolen (2005) otorgan énfasis a las consecuencias para el profesorado y el estudiantado, particularmente el efecto que podría tener la retroalimentación, las técnicas y los instrumentos administrados en la motivación y los hábitos de estudio.

Bonnen (2013), Gipps (1994), Medina Gual (2013), McMillan (1997) y Muñoz y Fonseca-Pedrero (2008) incluyen, además, una serie de propuestas para evidenciar la validez de las interpretaciones de la información recopilada, así como la confiabilidad. Suskie (2006, p. 37) plantea cuatro características de los instrumentos útiles (produzcan información certera y precisa, tengan un propósito claro, involucren al profesorado y al estudiantado, y se enfoquen en metas claras e importantes del aprendizaje estudiantil). Medina Gual (2013) propone un esquema con evidencias curriculares, interpretativas e instrumentales y sus respectivas estrategias. Gibbs (1994, p. 174) presenta seis criterios de calidad de los instrumentos: (a) fidelidad curricular, (b) comparabilidad, (c) dependencia, (d) credibilidad pública, (e) descripción del contexto y (f) equidad. De éstos, destacamos la equidad en el trato justo al estudiantado, proveyendo varias oportunidades y los acomodos necesarios para alcanzar los aprendizajes espera-

dos. Por su parte, McMillan (1997, p.49) plantea los siguientes: (a) objetivos de aprendizaje claros y apropiados, (b) métodos de evaluación apropiados, (c) validez, (d) confiabilidad, (e) justicia, (f) consecuencias positivas, y (g) practicalidad y eficiencia. Estos últimos criterios incluyen varios aspectos que merecen consideración del profesorado, tales como el tiempo y los recursos que consume preparar el instrumento, administrarlo y calificar la ejecución del estudiantado, así como la complejidad de la interpretación de los resultados. Por ejemplo, la construcción de una prueba objetiva toma mucho tiempo, pero conlleva poco tiempo contestarla y corregirla. Una tarea de ejecución requiere mayor cantidad de tiempo en elaborarla, contestarla y valorar el desempeño. Por esta razón, recomendamos que se revisen y se vuelvan a usar algunos ítems o tareas, cada cierto tiempo si es posible.

3.2. Confiabilidad

La confiabilidad o fiabilidad se refiere a la precisión o consistencia de las puntuaciones o de la información conseguida con un instrumento administrado en varias ocasiones. También está atada a la precisión o exactitud de las puntuaciones u otra información de un grupo de estudiantes, con los menores errores posibles. Estos errores podrían estar vinculados con los cambios en las condiciones de administración del instrumento, la subjetividad en la corrección o calificación del profesorado, la ambigüedad en los ítems, así como a la falta de motivación y la adivinanza del estudiantado.

La cantidad de ítems, tareas y momentos en que se utilicen los instrumentos es un factor relacionado con la confiabilidad. Por lo general, aumentar la cantidad de ítems en una prueba objetiva, incrementa el coeficiente de confiabilidad. Hay tres tipos principales de coeficientes de confiabilidad que se obtienen mediante procedimientos estadísticos: (a) estabilidad (o test-retest), que se refiere a la consistencia de las puntuaciones a lo largo del tiempo o en diferen-



tes ocasiones; (b) equivalencia, el cual concierne a determinar si dos o más formas paralelas de un instrumento producen puntuaciones o resultados similares; y (c) consistencia interna, que se enfoca en la cohesión de las respuestas a los ítems de un instrumento, que intenta medir o representar el mismo objetivo o contenido. Teóricamente, si un instrumento produce puntuaciones confiables estas deben ser similares para el grupo de estudiantes que lo contesta en dos o más ocasiones. La correlación entre las puntuaciones es el coeficiente de estabilidad.

En la mayoría de las aulas es imposible llevar a cabo una doble administración, así que se recurre a la confiabilidad basada en la consistencia interna, la cual requiere solo una. Una prueba objetiva cuenta con una clave para corregir las respuestas a los ítems, por lo cual la subjetividad en la corrección no es una limitación. El profesor o la profesora que la administra a grupos grandes de estudiantes y cuenta con el equipo y el programado de computadoras para corregirlas, podría analizar las respuestas a los ítems y calcular un coeficiente de confiabilidad (por lo general, de consistencia interna). Esto si las puntuaciones se van a interpretar con referencia a un grupo normativo. También, podría estimar el error estándar de medición, si lo desea. Este indica la precisión en las puntuaciones individuales del instrumento y depende de la magnitud del coeficiente de confiabilidad y la variabilidad de las puntuaciones; esto es, que a mayor confiabilidad menor es el error de medición. A menudo, estas estadísticas no se calculan en las pruebas que administran en los cursos o departamentos universitarios. Sin embargo, tomarlas en consideración para revisar la calidad técnica es un esfuerzo meritorio, máxime cuando se proclama que “es una prueba confiable”.

Con respecto a las tareas de ejecución (e.g., proyecto, portafolio) imprescindibles en las metodologías activas, la subjetividad en la calificación se reduce, pero no se extingue, con el uso de una rúbrica, lista de cotejo o escala de categorías que contenga los criterios e indicado-

res apropiados (Medina-Díaz & Verdejo-Carrión, 2019; Reddy & Andrade, 2010; Selke, 2013; Van der Schaaf, Baartman & Prins, 2012). Además, el estudiantado debe conocer éstos de antemano o puede participar en su confección. Si es posible, se aconseja presentar ejemplos de respuestas esperadas, acciones o trabajos en los distintos niveles de ejecución, que la rúbrica incluya. En ausencia de estos, se describen los procedimientos para corregir o calificar las respuestas o los trabajos. Si se administran preguntas para elaborar la respuesta, se califican las respuestas de todo el grupo de estudiantes a una pregunta, antes de revisar las contestaciones de otra. Esto no solo ayuda a mantener consistencia en la calificación, sino a proveer retroalimentación al estudiantado cuando se discuten las respuestas a las preguntas.

De esta manera, la confiabilidad se manifiesta mediante la consistencia en la asignación de puntuaciones con una escala de valoración o una rúbrica para calificar la ejecución de cada estudiante en la tarea de ejecución en cuestión. Esto conlleva dos procedimientos para hallar la consistencia: Intra-jueces y extra-jueces. La consistencia intra-jueces depende de cómo el profesor o la profesora aplique determinado instrumento (e.g., una rúbrica), de manera estable, para calificar las respuestas o los trabajos del estudiantado. Para esto, puede volver a valorar una muestra de trabajos, previamente revisados y hallar la coincidencia en las puntuaciones otorgadas, así como identificar las discrepancias (Cizek, 2009). El acuerdo entre-jueces requiere que dos o más personas revisen y califiquen la ejecución. En las aulas, esto es inusual, a menos que se cuente con la colaboración de otro profesor o profesora o del estudiantado. Esta es una buena oportunidad para propiciar la participación del estudiantado como observadores o jueces, a modo de coevaluación. El profesor o la profesora y uno o más estudiantes, de manera independiente, califican la ejecución de un o una estudiante, usando el mismo instrumento. Luego, se comparan las puntuaciones y se calcula un porcentaje de acuerdo u otra estadística (Stemler, 2004). Los desacuerdos en las puntua-



ciones que sugieren poca o ninguna confiabilidad en las puntuaciones, pueden ser provocados por aplicar el instrumento de manera distinta o por posibles sesgos en la persona que califica (e.g., leniencia o severidad). Esto es otra ocasión para involucrar al estudiantado en el diálogo y la experiencia de calificar y evaluar, así como para entender la naturaleza del proceso y las apreciaciones comunes y discordantes en la ejecución. Además, el profesorado también se beneficia al confirmar lo acertado de las puntuaciones adjudicadas.

Por otra parte, Brookhart (2003) y Smith (2003) proponen visualizar la confiabilidad como suficiencia o abundancia de la información. Brookhart (2003, p.11) alude a la estabilidad de la información para detectar la diferencia entre lo esperado y el estado actual de la ejecución del estudiantado o la cantidad de información. Smith (2003, p.31) se refiere a conseguir información suficiente, para tener una visión completa del estudiante, y que conduzca a una buena decisión. Asimismo, hay que considerar si el estudiantado tuvo varias oportunidades (ítems o tareas) y momentos para mostrar sus aprendizajes y lo que es capaz de hacer. Esto permite observar la variación y la consistencia en la ejecución y así, formular interpretaciones más acertadas de los aprendizajes logrados. Un profesor o una profesora podrían utilizar esta información y otra complementaria (e.g., entrevista) para derivar inferencias acerca de los aprendizajes logrados por el estudiantado. Además, ayuda a reducir la ansiedad o temor que podría provocar si solo cuenta con una ocasión o instrumento para demostrar lo aprendido. Como se aprecia, en estos casos no se calcula, ni es necesario, un coeficiente de confiabilidad. La confiabilidad depende del uso de varias técnicas que arrojen información consistente de los aprendizajes esperados. Cabe recalcar que tener puntuaciones o información confiable no es suficiente para declarar la validez.

Por último, la conveniencia de utilizar varios instrumentos para recoger información permite superar las limitaciones de cada uno y

tratar de representar la complejidad y las múltiples dimensiones del aprendizaje estudiantil. La combinación de la información cuantitativa y cualitativa acumulada a lo largo del proceso de enseñanza-aprendizaje ofrece una mirada más integral y precisa de los aprendizajes desarrollados y así, permite tomar decisiones y formular juicios adecuados y justos. Si se consideran puntuaciones, tales como las de una prueba o de la rúbrica aplicada a una tarea de ejecución, las medidas de tendencia central y variabilidad son útiles para describir la ejecución de un grupo. Estas también se comparten con el estudiantado. Cuando hay información de índole cualitativa (e.g., ensayo, reflexión) o gráfica (e.g. tirilla cómica, infograma), se consideran ciertos criterios (e.g., vocabulario, argumentación, uso de ejemplos y símbolos) para describir el desempeño o desarrollo de cada estudiante. Aquí se pueden utilizar estrategias de análisis para identificar elementos o patrones comunes o divergentes en las piezas escritas o gráficas. La información recopilada con varios instrumentos se agrega y compara (o “triangula”) para identificar temas o patrones de ejecución referentes a los aprendizajes, que convergen en decisiones y juicios acerca del estudiantado y el proceso instruccional: ¿Qué aprendizajes lograron? ¿Qué dificultades presentaron? ¿Qué se infiere acerca de los aprendizajes? ¿Cómo se van a usar los resultados? ¿Qué cambios se necesitan en las estrategias de enseñanza o en los instrumentos?

4. Discusión y conclusiones

La evaluación del aprendizaje constituye un proceso sistemático de emitir un juicio fundamentado en la información recopilada acerca los aprendizajes desarrollados y logrados, a lo largo del proceso de enseñanza-aprendizaje. La confianza depositada en la información recopilada depende la calidad de los instrumentos que el profesorado construye, administra y utiliza. La validez radica en lo apropiado y creíble de las interpretaciones de las puntuaciones o la infor-



mación recopilada un instrumento acerca de los aprendizajes de un o una estudiante. La evidencia recopilada mediante las distintas fuentes afianza la certeza de las interpretaciones y las inferencias, tanto del proceso de aprendizaje como del aprovechamiento logrado por el estudiantado. Además, ayuda a darle sentido a las decisiones que se toman en el proceso de enseñanza-aprendizaje (e.g., ampliar la discusión de un tema, ofrecer una experiencia práctica, utilizar un método activo en la enseñanza o recomendar tutorías) y a los juicios emitidos (e.g., “María ha logrado las competencias”). Estas decisiones y juicios, sin duda, también merecen ponderación.

La validez y la confiabilidad, como apun-
tamos, no han tenido suficiente atención en la discusión de las técnicas y los instrumentos de recopilar información, aplicados por el profesorado para evaluar el aprendizaje estudiantil. Posiblemente, aparezcan en una “lista de las especies en peligro de extinción” como comenta Popham (2005, p.71). Esperamos que este escrito contribuya a la reflexión y acción para preservarlos. Exhortamos a tener mayor prudencia en las interpretaciones de la información adquirida con los instrumentos que se elaboren y utilicen junto con las metodologías activas de enseñanza, si no hay evidencia de la validez que las respalde. Recomendamos, al menos, que se tome en cuenta la evidencia relacionada con el contenido y la consistencia de las puntuaciones o de la información obtenida, al emitir juicios y tomar de decisiones que afectan al estudiantado. Reconocemos, sin embargo, lo trabajoso y ambicioso de esta solicitud, ante la realidad de la labor docente y el respeto a la libertad académica que cobija al profesorado en las distintas instituciones universitarias. Frente a la posible fragilidad de las interpretaciones y decisiones derivadas de la información obtenida o acumulada, mayor evidencia apremia para respaldarlas.

Por supuesto, no basta con que el profesorado construya y administre mejores instrumentos, si no utiliza la información de manera apropiada, consistente y justa en la evaluación

del aprendizaje estudiantil. Como advierten Palomba y Banta (1999) y Banta y Pike (2012), se necesita usar los resultados para cerrar el ciclo de la evaluación (planificar-recoger información-interpretarla-usar los resultados). Tampoco es suficiente la información referente a los aprendizajes cognoscitivos, si no se contemplan los afectivos, psicomotores, sociales y otros que son relevantes en las distintas disciplinas universitarias (e.g., seguridad en el manejo de materiales o sustancias, autonomía y trabajo en equipo). Además, la evaluación del aprendizaje, de manera ética y constructiva, apunta al derecho del estudiantado a recibir notificación acerca del proceso evaluativo y a participar del mismo, de los criterios, las técnicas y los instrumentos que se emplean; así como de la interpretación de la información recopilada, tanto para mejorar el aprendizaje como la enseñanza. Hay que recordar que las interpretaciones, las inferencias y decisiones tomadas tienen consecuencias (unas más graves que otras), para el estudiantado y la sociedad. Finalmente, la expectativa de mejorar el aprendizaje estudiantil no depende exclusivamente de la evaluación, si no ocurren cambios en la visión del aprendizaje, las estrategias de enseñanza, el currículo de las disciplinas, el desarrollo profesional y el liderazgo del profesorado y la colaboración de la administración de las universidades para apoyarla, integrarla y valorarla.

Referencias bibliográficas

- Alquraan, M.F. (2012). Methods of assessing students' learning in higher education An analysis of Jordanian college and grading system. *Education, Business and Society: Contemporary Middle Eastern Issues*, 5(2), 124-133.
<https://doi.org/10.1108/17537981211251160>
- Andreu-Andrés, M.A., & Labrador-Piquer, M.J. (2011). Formación del profesorado en metodologías y evaluación. Análisis cualitativo. *Revista de Investigación en Educación*, 9(2), 236-245. <https://bit.ly/2UhFxu0>



- Angelo, T.A., & Cross, K.P. (1993). *Classroom assessment techniques* (2da ed.). Jossey-Bass.
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education (2018). *Estándares para pruebas educativas y psicológicas* (M. Lieve, Trans.). American Educational Research Association (Original work published 2014).
- Baartman, L.K.J., Bastianens, T.J., Kirschner, P.A., & Van der Vieuten, C.P.M. (2006). The wheel of competency assessment: Presenting quality criteria for competency assessment programs. *Studies in Educational Evaluation*, 32 (2), 153-170.
<https://doi.org/10.1016/j.studuc.2006.04.006>
- Banta, T.W., Lund, J.P., Black, K.E., & Oblander, F.W. (1996). *Assessment in practice: Putting principles to work on college campuses*. Jossey-Bass.
- Banta, T.W., & Pike, B.R. (2012). The bottom line: Will faculty use assessment findings? In C. Secolsky & B. Denison (Eds.), *Handbook on measurement, assessment and evaluation in higher education* (pp. 47-56). Routledge.
- Barkley, E.F., & Major, C.H. (2016). *Learning assessment techniques: A Handbook for college faculty*. Jossey-Bass.
- Black, P., & William, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
<https://doi.org/10.1080/0969595980050102>
- Black, P., & William, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80, 139-144.
- Bearman, M., Dawson, P., Bennett, S., Hall, M., Molloy, E., Boud, D. & Joughin, G. (2017). How university teachers design assessments: A cross disciplinary study. *Higher Education*, 74 (1), 49-64.
<https://doi.org/10.1007/s10734-016-0027-7>
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18, 5-25.
<https://doi.org/10.1080/0969594X.2010.513678>
- Bonnen, S.M. (2013). Validity in classroom assessment: Purposes, properties, and principles. In J.H. McMillan (Ed.), *Sage Handbook of Research on Classroom Assessment* (pp. 87-106). Sage.
- Brookhart, S.M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5-12.
<https://doi.org/10.1111/j.1745-3992.2003.tb00139.x>
- Brookhart, S.M. (2004). Assessment theory for college classrooms. *New Directions for Teaching and Learning*, 100, 5-14.
<https://doi.org/10.1002/tl.165>
- Brookhart, S.M. (2007). Expanding views about formative classroom assessment: A review of the literature. In J.H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 43-62). Teacher College Press.
- Brown, G.T.L. (2010). The validity of examination essays in higher education: Issues and responses *Higher Education Quarterly*, 64(3), 276-291.
<https://doi.org/10.1111/j.1468-2273.2010.00460.x>
- Brown, G., & Atkins, M. (1988). *Effective teaching in higher education*. Routledge.
- Cizek, G.J. (2009). Reliability and validity of information about student achievement: Comparing large-scale and classroom testing contexts. *Theory into Practice*, 48(1), 63-71.
<https://doi.org/10.1080/00405840802577627>
- Cizek, G.J. (2015). Validating test score and detecting test score use: Different aims, different methods. *Assessment in Education: Principle, Policy & Practice*, 23(2), 212-225.
<https://doi.org/10.1080/0969594X.2015.1063479>
- Davies, M.S., & Taras, M. (2018). Coherence and disparity in assessment literacies among higher education staff. *London Review of Education*, 16(3), 474-490.
<https://doi.org/10.18546/LRE.16.3.09>
- De la Orden Hoz, A. (2011). Reflexiones en torno a las competencias como objeto de evaluación en el ámbito educativo. *Revista Electrónica de Investigación Educativa*, 13(2), 1-21.
<https://bit.ly/2QIILUX>
- Epstein, R.M. (2007). Assessment in medical education. *New England Journal of Medicine*, 356 (4), 387-396.
<https://doi.org/10.1056/NEJMra054784>
- Erwin, T.D. (1991). *Assessing student learning and development*. Jossey-Bass.
- Esteve Zarazaga, J.M. (2007). Un examen a la cultura escolar. *Avances en Supervisión Educativa*, (7). <https://bit.ly/3ajVMfN>



- Fernández March, A. (2006). Metodologías activas para la formación de competencias. *Educatio Siglo XXI*, 24, 36-56. <https://bit.ly/39deN20>
- García-Merino, J. D., Urionabarrenetxea, S., & Bañales-Mallo, A. (2016). Cambios en metodologías docente y de evaluación: ¿Mejoran el rendimiento del alumnado universitario? *Revista Electrónica de Investigación Educativa*, 18(3), 1-18. <https://bit.ly/2WII8yE>
- Gil Flores, J. (2005). Valoraciones del alumnado universitario sobre las pruebas objetivas. *Revista de Investigación Educativa*, 23(1), 259-277. <https://bit.ly/2wE6z5B>
- Gilles, J.L., Detroz, P., & Blais, J.G. (2011). An international online survey of the practices and perceptions of higher education professors with respect to the assessment of learning in the classroom. *Assessment & Evaluation in Higher Education*, 36(6), 719-733. <https://doi.org/10.1080/02602938.2010.484880>
- Gipps, C.V. (1994). *Beyond testing: Towards a theory of educational assessment*. Falmer Press.
- Goñi-Zabala, J.M. (2005). *El espacio europeo de educación superior, un reto para la universidad*. Ediciones Octaedro.
- Goubeaud, K. (2010). How is science learning assessed at the postsecondary level? Assessment and grading practices in college Biology, Chemistry and Physics. *Journal of Science Education and Technology*, 19 (3), 237-245. <https://doi.org/10.1007/s10956-009-9196-9>
- Goubeaud, K., & Yan, W. (2004). Teachers educators teaching methods, assessments, and grading: A comparison of higher education faculty's instructional practices. *The Teacher Educator*, 40 (1), 1-16. <https://doi.org/10.1080/08878730409555348>
- Haladyna, T.M. (1997). *Writing test items to evaluate higher order thinking*. Allyn & Bacon.
- Hernández, R. (2012). Does continuous assessment in higher education support student learning? *Higher Education*, 64(4), 489-502. <https://doi.org/10.1007/s10734-012-9506-7>
- Huba, M.E., & Freed, J.E. (2000). *Learned-centered assessment on college campuses*. Allyn & Bacon.
- Hortigüela-Alcalá, D., Pérez-Pueyo, A., & López-Pastor, V. (2015). Implicación y regulación del trabajo del alumnado en los sistemas de evaluación formativa en educación superior. *Revista Electrónica de Investigación y Evaluación Educativa*, 21(1), ME6. <https://doi.org/10.7203/relieve.21.1.5171>
- Jacobs, L.C., & Chase, C.I. (1992). *Developing and using tests effectively: A guide to faculty*. Jossey-Bass.
- Joint Committee on Standards for Educational Evaluation (2015). *Classroom assessment standards: Practices for PreK-12 teachers*. Sage.
- Krzykowski, L., & Kinser, K. (2014). Transparency in student learning assessment. *Change*, 46(3), 67-73. <https://doi.org/10.1080/00091383.2014.905428>
- López-Pastor, V., & Sicilia-Camacho, A. (2016). Formative and shared assessment in higher education. Lessons learned and challenges for the future. *Assessment & Evaluation in Higher Education*, 42 (1), 77-97. <http://dx.doi.org/10.1080/02602938.2015.1083535>
- Mateo, J., & Martínez, F. (2008). *Medición y evaluación educativa*. La Muralla.
- McClarty, K.L., & Gaertner, M.N. (2015). *Measuring mastery: Best practices for assessment in competency-based education*. American Enterprise Institute for Public Policy Research. <https://bit.ly/2MN1m04>
- McMillan, J.H. (1997). *Classroom assessment: Principles and practice for effective instruction*. Allyn and Bacon.
- McMillan, J.H. (2008). *Assessment essentials for standard-based education* (2da ed.). Corwin Press.
- Medina-Díaz, M., & Verdejo-Carrión, A.L. (2019). *Evaluación del aprendizaje estudiantil* (6ta ed.). Autoras.
- Medina-Gual, L. (2013). La evaluación en el aula: Reflexiones sobre sus propósitos, validez y confiabilidad. *Revista Electrónica de Investigación Educativa*, 15(2), 34-50. <https://bit.ly/2xm8HiH>
- Messick, S. (1989). Validity. En R.L. Linn (Ed.), *Educational measurement* (3ra ed., pp. 13-103). Macmillan.
- Moss, P.A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice*, 22(4), 13-25. <https://doi.org/10.1111/j.1745-3992.2003.tb00140.x>
- Muñiz, J., & Fonseca-Pedrero, E. (2008). Construcción de instrumentos de medida para la evaluación



- ción universitaria. *Revista de Investigación en Educación*, (5), 13-25. <https://bit.ly/39izB8f>
- Newble, D., & Cannon, R. (1991). *A handbook for teachers in university and colleges*. Kogan Page.
- Nitko, A.J., & Brookhart, S.M. (2011). *Educational assessment of students* (6ta ed.). Pearson.
- Olmos-Miguelañez, S., & Rodríguez-Conde, M.J. (2010). Diseño del proceso de evaluación de los estudiantes universitarios españoles: ¿Respondiendo a una evaluación por competencias en el espacio europeo y Educación Superior? *Revista Iberoamericana de Educación*, 53(1). <https://bit.ly/2JsJM4B>
- Palomba, C.A., & Banta, T.W. (1999). *Assessment essentials: Planning, implementing, and improving assessment in higher education*. Jossey-Bass.
- Pereira, D.R., & Flores, M.A. (2016). Conceptions and practices of assessment in higher education: A study of Portuguese university teachers. *Revista Iberoamericana de Evaluación Educativa*, 9(1), 9-29. <http://dx.doi.org/10.15366/rie2016.9.1.001>
- Popham, W.J. (2005). *Classroom assessment: What teachers need to know* (4ta ed.). Pearson.
- Poskanzer, S.G. (2002). *Higher education law: The faculty*. John Hopkins University Press.
- Pozo, J.I., & Pérez-Echevarría, M. del R. (2009). (Coords.), *Psicología del aprendizaje universitario: La formación de competencias*. Morata.
- Rawlasyk, P.E. (2018). Assessment in higher education and student learning. *Journal of Instructional Pedagogies*, 21, 1-34. <https://bit.ly/2vPtbtT>
- Reddy, Y.M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435-448. <https://doi.org/10.1080/02602930902862859>
- Ricoy, M.C., & Fernández-Rodríguez, J. (2013). La percepción que tienen los estudiantes universitarios sobre la evaluación: Un estudio de caso. *Educación XXI*, 16(2), 321-342. <https://doi.org/10.5944/educxx1.2.16.10344>.
- Selke, M.J.G. (2013). *Rubric assessment goes to college*. Rowman & Littlefield.
- Smith, J.K. (2003). Reconsidering reliability in classroom assessment and grading. *Educational Measurement: Issues and Practice*, 22(4), 26-33. <https://doi.org/10.1111/j.1745-3992.2003.tb00141.x>
- Stemler, S.E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, 9, Article 4. <https://doi.org/10.7275/96jp-xz07>
- Stiggins, R. (2017). *The perfect assessment system*. Association for Supervision and Curriculum Development.
- Suskie, L. (2009). *Assessing student learning: A common sense guide* (2da ed.). Jossey-Bass.
- Taylor, C.S., & Nolen, S. B. (2005). *Classroom assessment: Supporting teaching and learning in real classrooms*. Pearson.
- UNESCO (1998). *La educación superior en el siglo XXI: Visión y acción*. Conferencia mundial sobre la Educación Superior, Paris. <https://bit.ly/33Qb4GG>
- Van der Schaaf, M., Baartman, L., & Prins, F. (2012). Exploring the role of assessment criteria during teachers' collaborative judgement processes of students' portfolios. *Assessment & Evaluation in Higher Education*, 37(7), 847-860. <https://doi.org/10.1080/02602938.2011.576312>
- Voorhees, R.A. (2001). Competency-based learning models: A necessary future. *New Directions for Institutional Research*, 110, 5-13. <https://doi.org/10.1002/ir.7>
- Webber, K. (2012). The use of learner-centered assessment in US Colleges and Universities. *Research in Higher Education*, 53(2), 201-228. <https://doi.org/10.1007/s11162-01119245-0>.
- Weimer, M. (Ed.). (2013). *Grading strategies for the college classroom: A collection of articles for faculty*. Magna Publications.
- Wolf, K., Dunlap, J., & Stevens, E. (2012). Ten things every professor should know about assessment. *The Journal of Effective Teaching*, 12(2), 65-79. <https://bit.ly/2UBD6Bh>
- Yükselii, H.S., & Gündüz, N. (2017). Formative and summative assessment in higher education: Opinions and practices of instructors. *European Journal of Education Studies*, 3(8), 336-356. <https://doi.org/10.5281/zenodo.832999>

