



Nº Extraordinario (Julio, 2017)

ISSN 1138-414X, ISSNe 1989-639X

Fecha de recepción: 25/01/2016

Fecha de aceptación: 05/07/2016

## **DISEÑO Y VALIDACIÓN DE LA ESCALA EVALUACIÓN DE LOS APRENDIZAJES (EEA)**

*Design and validation of the Assessment of Learning Scale (EEA)*



Melany Hebles Ortiz, Manuel Alonso-Dos-Santos,  
Concepción Yaniz Alvarez-de-Eulate y

Lourdes Villardon-Gallego

Universidad Católica de la Santísima Concepción  
(UCSC) CHILE

E-mail: [mhebles@ucsc.cl](mailto:mhebles@ucsc.cl)

### **Resumen:**

*El objetivo de esta investigación ha sido diseñar y validar un instrumento para evaluar la práctica evaluadora en la enseñanza universitaria desde un enfoque de Evaluación Orientada al Aprendizaje, caracterizado por incorporar las actividades de evaluación como actividades de aprendizaje, implicar a los estudiantes en el proceso y ofrecer retroalimentación con información precisa para orientar cambios hacia la mejora del aprendizaje. Para ello se elaboró la Escala sobre la Evaluación de los Aprendizajes (EEA) y se aplicó a una muestra de 1205 estudiantes de una Universidad Chilena. A partir de los datos recogidos se evaluaron las propiedades psicométricas de la escala propuesta así como su capacidad para reflejar los conceptos teóricos y constructos utilizados. Para el análisis estadístico se empleó el método de mínimos cuadrados parciales (PLS, partial least squares) en un modelo de ecuaciones estructurales (SEM, structural equation modeling). Los resultados indican que la escala es válida y fiable para medir el proceso de evaluación de los aprendizajes. Se han corroborado la fiabilidad de los ítems y de las dimensiones, y su validez discriminante y convergente. En consecuencia, los resultados de la investigación permiten confirmar el modelo teórico de nueve dimensiones de la Evaluación Orientada al Aprendizaje: coherencia entre objetivos de aprendizaje y evaluación, cobertura, utilización de los resultados, planificación, contenidos, sistematización, agentes, retroalimentación y momentos.*

**Palabras clave:** Escala, evaluación, Universidad, evaluación orientada al aprendizaje, retroalimentación, validación.

**Abstract:**

*The goal of this research has been to design and validate an instrument to evaluate the assessment process in higher education from the perspective of a learning-oriented assessment, characterized by incorporating evaluation activities as learning activities, involve students in the process and provide accurate feedback to guide changes towards improving learning. For this, a scale on the Assessment of Learning (EEA, Escala de Evaluación de los Aprendizajes) was developed and applied to a sample of 1205 students of a Chilean university. The psychometric properties of the proposed scale were evaluated as well as its capability to reflect the theoretical concepts and constructs used. For statistical analysis, the method of partial least squares (PLS) was used in a structural equation modeling (SEM). The results show that the scale is valid and reliable to measure the process of learning assessment. Moreover, they have confirmed the reliability of the items, dimensions and discriminant and convergent validity. Consequently, the results of the research can confirm the theoretical model of nine dimensions of learning-oriented assessment: coherence between learning objectives and evaluation, hedging, use of results obtained, planning, contents, systemizing, agents, feedback and moments.*

**Key Words:** *Assessment, feedback, higher education, learning oriented assessment, scale, validation.*

## 1. Introducción

La importancia de la evaluación dentro de los procesos de aprendizaje es ampliamente reconocida desde mediados del siglo XX (Wiliam, 2011). La evaluación es una actividad compleja considerada un pilar fundamental de la enseñanza universitaria (Álvarez Valdivia, 2008) y un elemento clave del proceso de aprendizaje (Villardón, 2006). La enseñanza universitaria debe facilitar el pensamiento independiente y creador del estudiante, y promover la adquisición de una amplia variedad de competencias, pero no siempre se tienen en cuenta estas finalidades en la evaluación.

El discurso dominante en la Universidad equipara evaluación y calificación, está dirigida por el docente y proporciona escasa retroalimentación (Gil-Flores, 2012; Ibarra y Rodríguez, 2010). Diversas investigaciones han confirmado que se evalúa estereotipadamente, sin adaptarse a los contextos, momento o aspectos a evaluar y con instrumentos inadecuados -por ejemplo, utilizando respuestas a preguntas para evaluar habilidades prácticas-, y no se informa al alumno de las condiciones de la evaluación (Contreras, 2010; Ricoy y Fernández-Rodríguez, 2013; Santos Guerra, 1993).

Las propuestas para mejorar la evaluación se centran en aspectos técnicos, como la selección y aplicación de instrumentos adecuados, sin una reflexión que la sustente y adoleciendo de sentido (Broabfoot y Black, 2004). Por ello, es clave convertir la práctica evaluativa en una oportunidad de aprendizaje y generar propuestas para llevar a cabo una evaluación centrada en el aprendizaje más que en la medición (Carless, Joughin y Mok, 2006).

A la evaluación del aprendizaje se le han atribuido dos funciones principales, una acreditadora del logro de determinados resultados, denominada sumativa, y otra orientada a fortalecer el propio aprendizaje, denominada formativa. Diversos autores

proponen la supresión de tales distinciones y la incorporación de un enfoque de la evaluación integrado en los propios procesos de aprendizaje. Álvarez Valdivia (2008) analizó 275 artículos sobre evaluación concluyendo que es necesaria una nueva cultura de la evaluación en la que ésta forme parte del proceso de aprendizaje y se oriente a la mejora del mismo. La evaluación universitaria debe adoptar una modalidad de evaluación auténtica, en la que cambian las prácticas, pero sobre todo cambian las concepciones de docentes y estudiantes sobre este proceso. Esta perspectiva sitúa la evaluación en el centro del proceso educativo y con una orientación al aprendizaje (Fernández, 2010).

Con este enfoque, Carless, (2007) propone el término *Evaluación Orientada al Aprendizaje* (EOA), destacando el papel que tiene la evaluación en el aprendizaje más que en la medición. Esta debe fomentar el aprendizaje de forma permanente, contribuyendo al aprendizaje a lo largo de la vida (Carless et ál., 2006). Para lograrlo, se deben considerar una serie de características. En primer lugar, las tareas de evaluación deben tratarse como tareas de aprendizaje, diseñadas y alineadas con los objetivos de aprendizaje, basadas en situaciones reales y cercanas, que permita un aprendizaje significativo para el estudiante. En segundo lugar, la implicación de los estudiantes en la evaluación es clave para lo que es necesario aclarar a los estudiantes las metas de aprendizaje y los criterios que se aplicarán. Finalmente, es necesario dar una retroalimentación oportuna que permita a los estudiantes ir mejorando en el proceso.

### 1.1. Tareas de evaluación

Para la correcta integración de la evaluación, curriculum e instrucción (Knight, 2006) se necesita coherencia entre los objetivos de aprendizaje y los de evaluación (Villardón, 2006) y la valoración del aprendizaje con distintas formas (Murphy, 2006). Los estudios han intentado determinar si las prácticas de evaluación se encuentran alineadas con los enfoques de enseñanza y con lo que se ha determinado en el plan de estudios y la instrucción (Sheikh, 2013) y se han visto posibles incongruencias. Por ejemplo, las pruebas escritas son frecuentemente insuficientes para evaluar algunos aprendizajes e incongruentes con una enseñanza que pretende promover mayor actividad por parte del estudiante (Shepard, 2000). Biggs (2007) plantea que el aprendizaje de los estudiantes dependerá en gran parte de cómo ellos creen que serán evaluados; por tanto, las formas de evaluación influirán en el enfoque de aprendizaje. De esta manera, si se quiere asegurar un determinado resultado, debe existir alineación entre lo que se evalúa y el modo de hacerlo con los objetivos de aprendizaje propuestos. No obstante, con frecuencia, los docentes demuestran dificultades para asegurar la coherencia, ya que los aprendizajes descritos en los programas son muy amplios, ambiguos o insuficientes (Contreras, 2010).

Esta concepción de evaluación supone la valoración del aprendizaje a través de distintas formas; los tipos de instrumentos utilizados para recoger información y valorar los diferentes tipos de adquisiciones son un elemento fundamental para mantener la coherencia y garantizar la validez de la evaluación. Es necesario

seleccionar y utilizar diferentes instrumentos que permitan valorar, tanto adquisiciones parciales como conocimientos, habilidades y actitudes, así como su integración en competencias complejas. Hasta hace no mucho tiempo en la enseñanza universitaria se han considerado como contenidos de evaluación fundamentalmente los conocimientos adquiridos y se han valorado a través de pruebas escritas y orales (Villardón, 2006). Nuevas propuestas curriculares suponen nuevos enfoques de evaluación; sin embargo, los cambios no han sido sustanciales (Manrique Villavicencio, 2009). Las pruebas tradicionales, por ejemplo, han sido sustituidas por estrategias que intentan que los estudiantes recuperen información, sin poder evaluar diferentes habilidades del pensamiento de los estudiantes (Sheikh, 2013).

Es necesario, por tanto, un sistema de evaluación que permita favorecer el aprendizaje de saberes y recoger todos los resultados de aprendizaje requeridos (conocimientos, habilidades y actitudes) para dar respuesta a una situación determinada (Villardón, 2006). Para garantizar la presencia de estas características en los procesos de evaluación, debe existir un diseño adecuado de las tareas que permita distintas formas de evaluación coherentes con los objetivos pretendidos en la asignatura (Padilla Carmona y Gil Flores, 2008). El término evaluación incrustada ha sido usado por Wilson y Scalise (2006), para señalar que las actividades de evaluación deben estar asociadas a las actividades de aprendizaje, así, cualquier tarea que se utiliza para la evaluación es, al mismo tiempo, una actividad de aprendizaje, lo que está en la línea de la mencionada, ya que supone la alineación constructiva de objetivos, contenidos y evaluación (Carless, 2007). Dichas tareas deben estar basadas en situaciones auténticas, es decir, situaciones lo más cercanas a la realidad profesional a la que se enfrentarán los estudiantes, lo que hará aumentar la significación de los aprendizajes (Fernandez, 2010).

Por tanto, la selección de instrumentos de evaluación es relevante por sus implicaciones en el aprendizaje; bajo esta perspectiva, el docente puede utilizar diferentes técnicas, ya sean formales, las cuales presentan una estructura definida previamente, como exámenes tipo test y pruebas, técnicas semiformales, como ejercicios y talleres prácticos dentro o fuera de la clase, y/o técnicas informales, las cuales no requieren mayor preparación y son útiles para conocer de manera breve aspectos del proceso de aprendizaje, como por ejemplo, preguntas con el propósito de incentivar la reflexión de los temas revisados (Berliner, 1987). Aproximaciones más recientes han cuestionado la idoneidad de las técnicas que se han venido utilizando tradicionalmente en la Universidad. En los últimos años han aparecido nuevos métodos alternativos para dar respuesta a la evaluación de aspectos que trascienden lo conceptual, tales como portafolios, proyectos, debates, entre otras, para evaluar no sólo conocimientos, sino habilidades y actitudes (Padilla Carmona y Gil Flores, 2008).

## 1.2. Retroalimentación

La retroalimentación, eje fundamental de la evaluación formativa, tiene una función de mejora para el aprendizaje en la medida en que no sólo se le entrega al

estudiante información respecto a su desempeño, sino que también se le informa acerca del proceso seguido y de cómo puede mejorar en el futuro en tareas similares (Knight, 2006). No obstante, la retroalimentación ha sido particularmente inefectiva debido a que se entrega demasiado tarde como para que los estudiantes puedan utilizar la información productivamente (Carless, 2006). Un estudio realizado por Hernández (2012) demuestra que un alto porcentaje de estudiantes universitarios de primer año se muestran insatisfechos con la retroalimentación ya que los comentarios recibidos son limitados, orientados a la justificación de la calificación y no sugieren cómo mejorar el aprendizaje.

Gibbs y Simpson (2004) concluyen que la retroalimentación debe ser oportuna, es decir, los estudiantes deben recibirla en un momento en que puedan utilizarla para los fines del trabajo actual y también para realizar mejoras en el desarrollo de tareas futuras.

La retroalimentación no sólo tiene múltiples beneficios en la mejora de la tarea entregada por el estudiante, lo que se traduciría en mejores aprendizajes, también se relaciona directamente con la posibilidad de que éste se autoevalúe, y con una mayor autorregulación de su proceso de aprendizaje, lo que aumenta su confianza para participar en una tarea, ya que, al conocer cómo puede ir mejorando sus resultados, mejorará su sentido de autoeficacia (Hattie y Timperley, 2007).

A pesar de los beneficios de la retroalimentación, el uso de ésta en el aula es escasa, lo que produce que el estudiante aprenda de manera superficial y rutinaria (Hattie y Timperley, 2007).

### **1.3. Implicación de los estudiantes en la evaluación**

Otro componente esencial de la EOA es la implicación de los estudiantes en la evaluación, que supone participar como evaluadores de su propio aprendizaje. La incorporación de distintos agentes de evaluación a la práctica evaluativa, a través de la autoevaluación y la co-evaluación, permiten de manera directa la participación del estudiante en la evaluación (Padilla Carmona y Gil Flores, 2008). La autoevaluación favorece que el estudiante desarrolle su competencia reflexiva y de autocrítica, que tenga más claridad respecto a los criterios con los cuales se va a evaluar su trabajo pudiendo así realizar ajustes en aquello que se le solicita para mejorar la calidad y los resultados (Navarro Soria y González Gómez, 2010). La co-evaluación fomenta el aprendizaje tanto de los que evalúan el trabajo como de los que reciben las apreciaciones de los compañeros (Keppell, Au, Ma y Chan, 2006). En general ambas técnicas unidas a una adecuada retroalimentación por parte del profesor, favorecen el desarrollo de las habilidades metacognitivas de los estudiantes, que propician una toma de conciencia acerca de su propio proceso de aprendizaje, avances, estancamientos y acciones para progresar (Padilla Carmona y Gil Flores, 2008).

Para que los estudiantes sean partícipes de la evaluación es necesario que conozcan con claridad los criterios de evaluación que deben aplicar (Padilla Carmona y Gil Flores, 2008). Se propone que los estudiantes sean conscientes de los referentes

de los que parte un profesor a la hora de evaluar a sus estudiantes (Carless, 2006) a través de criterios claros y detallados (Ballantyne, Hughes, y Mylonas, 2002), los cuales deben ser comunicados a los estudiantes antes de realizar la evaluación (Tillema, Leenknecht y Segers, 2011). Así se facilita la igualdad de oportunidades a los estudiantes para demostrar sus competencias, lo que potencia la confianza de éstos hacia el sistema de evaluación (Van Gennip, Segers, y Tillema, 2009).

La capacidad de los docentes para comunicar los criterios con los que se evaluará y la siguiente comprensión de éstos por parte de los estudiantes constituye, por tanto, un aspecto central en una buena práctica de evaluación (Shepard, 2000); sin embargo, aparentemente es un aspecto difícil para los docentes, ya que la mayoría de las veces dichos criterios permanecen como un conocimiento tácito, sin que se hagan públicos ni se expresen concretamente (Contreras, 2010).

La comunicación de estos criterios, es sin duda, fruto de una adecuada planificación de la evaluación, otro aspecto fundamental, ya que se debe garantizar un diseño que permita mejorar el aprendizaje, valorando la implicación del estudiante en el proceso de aprendizaje (Sadler, 2010). Es importante que la planificación considere, además, los instrumentos de evaluación más adecuados para cubrir los logros pensados (Tillema et ál., 2011).

En este sentido hay que destacar que la EOA debe considerar una adecuada cobertura, la que se refiere a la amplitud de lo que se evalúa (Linn, Baker, y Dunbar, 1991). Teniendo en cuenta el impacto que lo evaluado tiene en el aprendizaje, una mayor cobertura se relacionará positivamente con una mayor calidad del aprendizaje. Por tanto, evaluar la totalidad de las adquisiciones previstas que habitualmente se incluyen en los programas formativos, como objetivos referidos a nuevos aprendizajes, resultados o competencias adquiridas es una característica de la EOA (Tillema et ál., 2011).

En complemento de lo anterior, en la EOA, se plantea necesario recoger información en distintos momentos del proceso de enseñanza-aprendizaje. Esto permite al docente evaluar tanto el proceso seguido como el producto de dichos aprendizajes. En la medida en que se realiza esta recogida y valoración a lo largo del desarrollo del aprendizaje, se puede proporcionar retroalimentación a los estudiantes en momentos adecuados para establecer las mejoras correspondientes y lograr que las mismas tengan consecuencias en la mejora de los resultados. Una evaluación completa al inicio, durante y al final del proceso es útil para múltiples propósitos vinculados con el aprendizaje, desde la identificación de los puntos débiles o conocimientos previos o la disposición de información oportuna a estudiantes durante el desarrollo de nuevas capacidades resultantes de su aprendizaje actual, hasta la identificación de estrategias para lograr futuros aprendizajes (Crisp, 2012).

La EOA es una práctica compleja que necesita planificación, una selección exhaustiva de los contenidos a evaluar y diversidad en las formas y momentos de evaluación; dichos elementos son condiciones entre otras, de la validez de dicha

evaluación. Dado que hay aspectos importantes a considerar al evaluar, que van más allá de los conocimientos que el alumno posee, como la responsabilidad, la creatividad, la actitud, etc., la evaluación debiera ser un proceso continuo que brinde información necesaria para la toma de decisiones en respuesta a los ritmos y estilos de aprendizajes de los estudiantes.

Segura (2007) añade que la evaluación, como una actividad sistemática y continua, presenta un carácter instrumental que debe ayudar al crecimiento personal de los estudiantes, valorando su rendimiento en torno a sus progresos con respecto a sí mismo, detectando las dificultades de aprendizaje y las deficiencias en el modo de enseñar para, en consecuencia, corregir o confirmar el currículo y los procedimientos pedagógicos utilizados.

Finalmente y en coherencia con lo anterior, como el foco de la EOA es la integración de la evaluación en la enseñanza de forma continua, se destaca la importancia de que los docentes utilicen los resultados de la evaluación para realizar ajustes en su enseñanza y, así, favorecer que los estudiantes mejoren su aprendizaje (Popham, 2008). De esta manera, el proceso de evaluación convierte al docente más que en un participante que certifica o aprueba, en un optimizador de los aprendizajes, que utiliza la información relevante para realizar modificaciones para mejorar su labor permitiendo que el que aprende tome conciencia de sí mismo y de sus metas a través de la reflexión y perfeccionamiento de su desempeño (Bordas y Cabrera, 2011).

Es decir, el enfoque de EOA concibe la evaluación como parte del proceso de aprendizaje, basada en criterios, de carácter colaborativo, que concede un papel protagonista a la retroalimentación como eje de la mejora del aprendizaje. Defiende la necesidad de estimular la participación de los estudiantes en la evaluación (Álvarez Valdivia, 2008), lo que motiva, mejora y consolida el aprendizaje de los estudiantes, favoreciendo así el aprendizaje a lo largo de la vida y para la vida (Boud y Falchikov, 2006).

En resumen, los elementos que caracterizan la EOA son los siguientes: coherencia entre los objetivos de aprendizaje y evaluación, esto es, utilización de tareas auténticas cercanas al desempeño profesional; una adecuada cobertura que permita valorar todos los resultados de aprendizaje pretendidos (contenidos, habilidades y actitudes); recogida de información en diferentes momentos de evaluación con funciones diagnósticas, formativas o acreditadoras; valoración del aprendizaje a través de técnicas de evaluación diversas, planificación de la evaluación e información al estudiante del sistema de evaluación establecido (técnicas, criterios, ...); retroalimentación, como eje fundamental del proceso de aprendizaje; implicación de los estudiantes en el proceso de evaluación, incorporando distintos agentes a la evaluación; y uso de los resultados de la evaluación para ajustar la enseñanza y el aprendizaje.

Son varios los estudios cuantitativos que han identificado los beneficios de la EOA, sin embargo, no se ha prestado especial atención a las propiedades

psicométricas de las escalas de medida que se utilizan para valorar de manera específica estas prácticas evaluadoras (Pat-El, Tillema, Segers, y Vedder, 2013). Generalmente, la evaluación ha sido estudiada como un ámbito o dimensión junto a otros en escalas de evaluación de la docencia.

Teniendo en cuenta esta realidad, el objetivo de esta investigación ha sido diseñar y validar un instrumento para evaluar la práctica evaluadora en la enseñanza universitaria desde un enfoque de Evaluación Orientada al Aprendizaje, atendiendo a los elementos y características descritas.

## 2. Método

El estudio ha consistido en el diseño y aplicación de la escala Evaluación de los Aprendizajes (EEA) a los estudiantes de la Universidad Católica de la Santísima Concepción (UCSC), (Concepción, Chile).

### 2.1 Muestra

La muestra la constituyeron 1205 estudiantes de pregrado de la UCSC. Se pretendió una representación proporcional de todas las carreras de pregrado, para lo cual primero se observó el número de estudiantes de cada carrera, luego se determinó la proporción que supone cada una de estas carreras respecto al total y finalmente se seleccionaron aleatoriamente de cada carrera los cursos y asignaturas para proceder con la recogida de datos, encuestando al azar tantos estudiantes de cada curso y asignatura como fuese necesario para cumplir con el porcentaje estimado. Se consideraron todas las facultades, y cada una de las carreras de la UCSC. De esta forma logramos tener el mayor grado de representatividad. Además con este tipo de muestreo el error muestral total nunca podrá ser mayor al mayor nivel de error muestral por facultad, fijado en el 5% y el nivel de confianza en el 95%.

Se encuestaron estudiantes de siete facultades, de los cuales 47% eran hombres y 53% mujeres. La edad media de los estudiantes era de 21.4 (D.T: 1.45), con un valor mínimo de 18 y máximo de 36. La muestra estuvo distribuida en cinco cursos, primer año con un 15.9%, segundo con un 29.1%, tercero con un 4.5%, cuarto un 28.1% y quinto un 22.4% del total de la muestra. La baja tasa de representación del tercer curso se debe al carácter práctico de éste que se realiza fuera de la Universidad. Respecto a la distribución de las facultades, aproximadamente un 10 % pertenecía a la Facultad de Ciencias Económicas y Administrativas, un 16.5% a la Facultad de Medicina, un 2.5% a la Facultad de Ciencias, un 8% a la Facultad de Derecho, un 25% a la Facultad de Ingeniería, 10% a la Facultad de Comunicación, Historia y Ciencias Sociales, y finalmente un 27% a la Facultad de Educación.

## 2.2. Instrumento

Para este estudio se diseñó una escala teniendo como base un modelo teórico de la Evaluación Orientada al Aprendizaje (Carless, 2007). Para la elaboración de la escala EEA se siguieron las siguientes etapas:

En la primera etapa se consideraron los artículos científicos que contuvieran algunas de las palabras en español y su respectiva traducción en inglés: Assessment for Learning, Learning Oriented Assessment, Assessment and Higher Education. Desde 1980 hasta hoy en las bases de datos 2014 para definir las dimensiones de una Evaluación Orientada al Aprendizaje y se seleccionaron aproximadamente cuatro ítems por dimensión, lo que derivó en un instrumento de 37 ítems distribuidos en nueve dimensiones: Objetos de Evaluación, Sistematización para Obtener Información, Utilización de los Resultados de la Evaluación, Articulación entre Objetivos y la Evaluación, Periodicidad de la Evaluación, Evaluación Diagnóstica, Planificación Evaluativa, Efectividad de la Evaluación y Retroalimentación.

En una segunda etapa se llevó a cabo un juicio de expertos. La consulta se realizó a cuatro profesores de dos universidades de la ciudad de Concepción (Chile) de las áreas de educación con experiencia en evaluación. Dichos expertos respondieron un cuestionario para la valoración de la pertinencia de los ítems en las dimensiones propuestas, la claridad de la formulación y la exhaustividad, es decir, la capacidad que tienen cada subconjunto de ítems para cubrir cada una de las nueve dimensiones abordadas, y de abarcar aspectos diferentes de una misma dimensión, con un espacio para hacer comentarios, y otro para indicar posibles solapamientos entre los ítems.

Del juicio de expertos se derivaron cambios en la escala. En primer lugar, se cambió la nomenclatura de la dimensión objetos de evaluación a cobertura de la evaluación, eliminando dos ítems (el 4 y el 7). Por recomendación de los expertos se elimina la dimensión efectividad de la evaluación (28, 29, 30 y 31), por considerarse difícil de valorar desde la perspectiva del alumno, ya que ésta más bien abordaba criterios de confiabilidad y validez de los instrumentos de evaluación. Los jueces sugieren que dos de los ítems de la dimensión sistematización deben constituir una nueva dimensión, denominada agentes de evaluación (20 y 21). Las variables periodicidad de la evaluación y evaluación diagnóstica fueron replanteadas en una sola dimensión denominada momentos de evaluación con tres ítems (25, 26 y 27). Asimismo, tras recomendación de expertos, se añade la dimensión contenidos de evaluación con tres ítems (13, 14 y 15).

La nueva escala presenta nueve dimensiones y 27 ítems: Coherencia entre Objetivos de Aprendizaje y Evaluación, Cobertura, Utilización de los Resultados, Planificación, Contenidos, Sistematización, Agentes, Retroalimentación y Momentos.

## 2.3. Procedimientos

Se contactó con los docentes para informarles de la investigación y para contar con la autorización para la aplicación colectiva del instrumento a los

estudiantes. Los estudiantes que respondieron fueron informados del objetivo y características de la investigación así como del carácter voluntario y anónimo de su participación. Aquellos que aceptaron participar firmaron su consentimiento. La recolección de datos fue realizada entre los meses mayo y junio de 2014 por estudiantes de la asignatura de Investigación de Mercados, para tal fin. El periodo de recolección de datos fue durante siete minutos en una hora lectiva.

Se procedió a evaluar las propiedades psicométricas de la escala propuesta así como su capacidad para reflejar los conceptos teóricos y constructos utilizados, a través de los datos procedentes de las observaciones. En el proceso de validación de la escala se siguen los criterios de Churchill (1979) para la construcción y desarrollo de mediciones en ciencias sociales, para la construcción de escalas de DeVellis (2003), y procedimientos y técnicas de validación de escalas de Mackenzie, Podsakoff y Podsakoff (2011).

Para el análisis estadístico se empleó el método de mínimos cuadrados parciales (PLS) en un modelo de ecuaciones estructurales (SEM). Se trata de una técnica multivariante que combina el análisis factorial con la regresión lineal con el objetivo de estimar el grado de ajuste de los datos muestrales a un modelo hipotético poblacional expresado mediante un diagrama de flujos. Para ello se empleó el paquete de software SmartPLS 2.0.M3. Las razones fundamentales que nos llevaron a emplear esta técnica son las siguientes: En primer lugar, PLS es una técnica SEM con base en el estudio de la varianza que se ha utilizado en la investigación previa para la construcción de escalas de medición (Howat y Assaker, 2013). Segundo, PLS no supone, a diferencia de los métodos basados en la covarianza, que la distribución sea normal multivariada y observaciones independientes (Castro, Carrión, y Salgueiro, 2007). En tercer lugar, se recomienda el uso de PLS cuando se abarcan cuestiones de investigación que han sido examinadas en la investigación previa, de manera limitada o el análisis es exploratorio (Petter, Straub, y Rai, 2007). Cuarto, cuando la estructura teórica es compleja, la probabilidad de obtener peores ajustes del modelo con métodos basados en la covarianza es mayor (Chin, 2010). Estas razones conllevan la elección de PLS como el método más adecuado para esta investigación.

Para una adecuada evaluación de las bondades de la escala debemos estimar la fiabilidad (de los ítems y dimensiones) y la validez (discriminante y convergente). En primer lugar, en general, mediante la fiabilidad se indica la precisión con la que un instrumento de medida arroja puntuaciones exentas de errores aleatorios (Peterson, 1994). En un modelo PLS, la fiabilidad individual del ítem se valora examinando las cargas ( $\lambda$ ), o correlaciones simples de los indicadores con su respectivo constructo. Se considera que un indicador que forma parte de un constructo ha de poseer una carga igual o superior a .707 lo que implica que la varianza compartida entre el constructo y sus indicadores es mayor que la varianza del error (Barclay, Higgins, y Thompson, 1995).

La fiabilidad de la dimensión evalúa la rigurosidad con la que los ítems están midiendo la misma variable latente, esto es, comprueba la consistencia interna. Uno

de los índices más comúnmente aceptados que permite estimar la homogeneidad de los ítems de medida, es el coeficiente alpha de Cronbach (Cronbach, 1951). El índice de fiabilidad compuesta, al igual que el alfa de Cronbach, es una medida de consistencia interna que evalúa la validez de los ítems para medir el concepto objeto de estudio. Nunnally (1978) sugiere .7 como nivel mínimo para una fiabilidad aceptable.

La validez se define como el grado en que las puntuaciones obtenidas con la escala reflejan las diferencias reales entre los objetos con respecto de la característica que se mide (Malhotra, 2004). Es decir, el grado en que un instrumento mide la característica que se está investigando. Debido a su complejidad, su asunción implica aceptar a su vez dos subtipos de validez: validez discriminante y validez convergente.

El procedimiento PLS, respecto a la validez discriminante mide el grado en que dos dimensiones diferentes se diferencian ciertamente. No deberíamos medir dos veces la misma característica. La evaluación de la validez discriminante en PLS requiere de dos métodos: (a) comparar si la varianza media compartida entre una dimensión y sus ítems es mayor que la varianza que dicha dimensión comparte con las otras dimensiones del modelo (Barclay et ál., 1995) y (b) analizar si las correlaciones entre las dimensiones son más bajas que la raíz cuadrada de la AVE (Fornell y Larcker, 1981). Por otra parte, la validez convergente, nos permite evaluar el grado en que los ítems de una dimensión explican la misma idea subyacente. Queremos comprobar que dos dimensiones que teóricamente deberían estar relacionadas, están, de hecho, relacionadas. Para ello se emplea el valor de la varianza extraída (AVE) que según Fornell y Larcker, (1981) debería superar el valor .5. De esta manera, más del 50% de la varianza de la dimensión sería debida a los ítems.

### 3. Resultados

En primer lugar se procedió al tratamiento de los valores perdidos según el método listwise (Little y Rubin, 1987). Se escogió por su fácil implementación y confección de estimadores válidos. El número de casos eliminados fue menor a 20, se trata de un porcentaje de casos muy pequeño respecto al total, considerados distribuidos al azar entre las facultades y cursos. Por lo tanto, sin repercusión en la investigación (Hair, 1999).

Los valores atípicos pueden influir en el análisis de datos, por ello es importante asegurarse de que las conclusiones no se basan en una o más observaciones extremas. Uno de los enfoques estadísticos comúnmente empleados para la identificación de los valores atípicos multivariantes se basa el uso de la distancia de Mahalanobis (D2). Este estadístico mide la distancia entre casos basándose en el centroide, teniendo en cuenta la covarianza de las variables en cuestión. Los valores D2 siguen una distribución chi-cuadrado con p grados de

libertad ( $p = \text{número de variables}$ ). Los valores D2 significativos al nivel .05 indican valores extremos. Entonces se eliminaron tres casos extremos de la base de datos.

Se procede a continuación a evaluar mediante PLS la escala propuesta. Primero se evalúan las cargas estandarizadas para valorar la fiabilidad de los ítems (Tabla I). Todas las cargas factoriales son superiores a .707. Y todas ellas son significativas con un valor de referencia de 0.01. Por lo tanto, la varianza compartida entre el constructo y sus indicadores es mayor que la varianza del error señalando que más del 50% de la varianza del parámetro observado es compartida con la dimensión (Fornell et ál, 1981).

Se emplea el análisis de fiabilidad compuesta para determinar la rigurosidad con la que los ítems están midiendo una misma dimensión. Para la fiabilidad compuesta, Nunnally (1978) sugiere valores adecuados a partir de .7. Para el coeficiente del alfa de Cronbach, Lévy, Mallou y González (2003) proponen un valor no menor a .8 en las fases de investigación concluyentes y .7 en las fases exploratorias. Para todas las dimensiones del modelo se superan los valores mínimos sugeridos para la fiabilidad compuesta. No así para el alfa de Cronbach con valores próximos a 0.7, en todo caso superiores a 0.6 (tabla 1).

Para examinar la validez convergente se ha empleado la varianza extraída de los constructos. El AVE proporciona la cantidad de varianza que un constructo obtiene de sus indicadores con relación a la cantidad de varianza debida al error de medida (Fornell y Larcker, 1981). Los citados autores recomiendan que el valor del AVE sea superior a 0.5. Todos los indicadores del índice muestran valores por encima de .055.

Los análisis de correlaciones cruzadas entre ítems y las dimensiones muestran los resultados de validez discriminante (Gefen, Straub, y Boudreau, 2000) (Tabla II y III). Para la primera comprobación se han tomado las correlaciones cruzadas y se han comprobado que no son significativas en relación con las cargas (Chin, 2010). Examinando la tabla 2 se comprueba que ningún ítem carga en otras dimensiones en mayor medida que en la suya propia. Lo que implica que el indicador en cuestión explica solo la dimensión que debe explicar diferenciándose de las demás.

Para el segundo método de análisis de la validez discriminante se ha comparado la raíz cuadrada de las AVE (los valores de la diagonal de la tabla 3) con las correlaciones de las variables latentes (parámetros por debajo de la diagonal). Todas las dimensiones se relacionan en mayor medida con sus propias medidas que con otras dimensiones. Es decir, las dimensiones son diferentes y reflejan un concepto subyacente diferente al resto de dimensiones del modelo.

Por tanto, los resultados expuestos indican que la escala es válida y fiable para medir el proceso de evaluación de los aprendizajes. Se han corroborado la fiabilidad de los ítems, de las dimensiones y su validez discriminante y convergente.

Tabla I

*Evaluación de fiabilidad y validez de la escala, análisis de la varianza (AVE), fiabilidad compuesta (CR), alfa de Cronbach (a) y cargas factoriales (B) con su significancia*

Código	Variables	AVE	CR	$\alpha$	Ítem	B
Cob	Cobertura de la evaluación	0.627	0.835	0.704	Obj1	0.778***
					Obj2	0.787***
					Obj3	0.810***
Sis	Sistematización para obtener información	0.552	0.872	0.731	Sis1	0.753***
					Sis2	0.734***
					Sis3	0.741***
					Sis4	0.742***
Uti	Utilización de los resultados de evaluación	0.692	0.871	0.780	Uti1	0.802***
					Uti2	0.859***
					Uti3	0.834***
Art	Coherencia entre objetivos y evaluación	0.717	0.884	0.804	Art1	0.824***
					Art2	0.877***
					Art3	0.838***
Pla	Planificación evaluativa	0.599	0.817	0.665	Pla1	0.753***
					Pla2	0.823***
					Pla3	0.742***
Ret	Retroalimentación	0.652	0.849	0.733	Ret1	0.781***
					Ret2	0.848***
					Ret3	0.793***
Age	Agentes de la evaluación	0.879	0.936	0.863	Age1	0.938***
					Age2	0.938***
Con	Contenidos de la evaluación	0.696	0.831	0.781	Con1	0.773***
					Con2	0.886***
					Con3	0.840***
Mom	Momentos de la evaluación	0.590	0.812	0.653	Mom1	0.792***
					Mom2	0.716***
					Mom3	0.794***

\*p < 0.10, \*\*p < 0.05, \*\*\*p < 0.01, \*\*\*\*p < 0.001

Fuente: Elaboración propia

Tabla II  
*Validez discriminante de los ítems, correlaciones cruzadas*

Item	Variable								
	Age	Art	Cob	Con	Mon	Pla	Ret	Sis	Uti
Age1	<b>0.938</b>	0.225	0.336	0.402	0.449	0.384	0.462	0.420	0.477
Age2	<b>0.938</b>	0.189	0.288	0.381	0.459	0.359	0.501	0.374	0.496
Art1	0.149	<b>0.824</b>	0.496	0.338	0.264	0.359	0.387	0.453	0.192
Art2	0.184	<b>0.877</b>	0.537	0.379	0.347	0.429	0.390	0.481	0.269
Art3	0.220	<b>0.838</b>	0.572	0.439	0.364	0.491	0.406	0.503	0.367
Cob1	0.201	0.563	<b>0.778</b>	0.391	0.396	0.443	0.405	0.452	0.353
Cob2	0.304	0.460	<b>0.787</b>	0.398	0.365	0.457	0.355	0.433	0.390
Cob3	0.284	0.490	<b>0.810</b>	0.404	0.410	0.449	0.405	0.443	0.419
Con1	0.438	0.337	0.378	<b>0.773</b>	0.470	0.536	0.429	0.484	0.443
Con2	0.299	0.417	0.455	<b>0.886</b>	0.475	0.498	0.424	0.506	0.364
Con3	0.330	0.390	0.420	<b>0.840</b>	0.444	0.464	0.444	0.525	0.365
Mom1	0.400	0.351	0.426	0.458	<b>0.792</b>	0.433	0.556	0.454	0.409
Mom2	0.278	0.340	0.370	0.398	<b>0.716</b>	0.396	0.411	0.446	0.276
Mom3	0.423	0.220	0.346	0.417	<b>0.794</b>	0.334	0.422	0.366	0.379
Pla1	0.280	0.367	0.452	0.433	0.382	<b>0.753</b>	0.362	0.422	0.501
Pla2	0.257	0.467	0.467	0.462	0.372	<b>0.823</b>	0.394	0.504	0.301
Pla3	0.401	0.339	0.391	0.489	0.415	<b>0.742</b>	0.460	0.444	0.400
Ret1	0.568	0.326	0.388	0.409	0.455	0.418	<b>0.781</b>	0.450	0.455
Ret2	0.370	0.392	0.414	0.466	0.518	0.435	<b>0.848</b>	0.452	0.428
Ret3	0.296	0.414	0.387	0.375	0.486	0.402	<b>0.793</b>	0.465	0.311
Sis1	0.415	0.440	0.473	0.583	0.485	0.503	0.499	<b>0.753</b>	0.410
Sis2	0.368	0.349	0.346	0.425	0.432	0.379	0.416	<b>0.734</b>	0.345
Sis3	0.177	0.454	0.401	0.333	0.344	0.412	0.354	<b>0.741</b>	0.217
Sis4	0.273	0.436	0.423	0.422	0.338	0.445	0.388	<b>0.742</b>	0.272
Uti1	0.411	0.267	0.395	0.332	0.356	0.330	0.362	0.289	<b>0.802</b>
Uti2	0.460	0.282	0.414	0.402	0.403	0.418	0.400	0.374	<b>0.859</b>
Uti3	0.424	0.283	0.414	0.409	0.400	0.510	0.461	0.389	<b>0.834</b>

Fuente: Elaboración propia

Tabla III  
 Validez discriminante entre constructos. Criterio de Fornell-Larcker

	Age	Art	Cob	Con	Mon	Pla	Ret	Sis	Uti
Age	<b>0.938</b>								
Art	0.220	<b>0.847</b>							
Cob	0.333	0.635	<b>0.792</b>						
Con	0.418	0.459	0.502	<b>0.834</b>					
Mon	0.484	0.389	0.493	0.552	<b>0.768</b>				
Pla	0.396	0.510	0.567	0.593	0.501	<b>0.774</b>			
Ret	0.513	0.466	0.491	0.517	0.602	0.519	<b>0.808</b>		
Sis	0.423	0.567	0.559	0.605	0.544	0.591	0.564	<b>0.743</b>	
Uti	0.519	0.334	0.491	0.462	0.466	0.513	0.495	0.426	<b>0.832</b>

Fuente: Elaboración propia

#### 4. Discusión y conclusiones

Los resultados de la investigación permiten confirmar las nueve dimensiones del modelo teórico planteado respecto a la Evaluación Orientada al Aprendizaje: Coherencia entre Objetivos de Aprendizaje y Evaluación, Cobertura, Utilización de los Resultados, Planificación, Contenidos, Sistematización, Agentes, Retroalimentación y Momentos.

Se ha diseñado una escala que demuestra ser estadística y conceptualmente fiable, consistente y válida para medir el proceso de evaluación de los aprendizajes desde una perspectiva de EOA. Todos los índices de fiabilidad, validez y consistencia se cumplieron siguiendo la metodología PLS. Una de las principales aportaciones del EEA, es que supera deficiencias de algunos instrumentos actuales como baja validez de constructo y consistencia interna de la escala, como es el caso del Assessment Experience Questionnaire propuesto por Gibbs y Simpson (2003). Esto se logra con la utilización de la técnica SEM, la cual no requiere distribuciones muestrales predeterminadas, muy poco frecuentes en Ciencias Sociales.

La escasez de escalas en español para medir procesos de evaluación de los aprendizajes desde los estudiantes ha impulsado la traducción al español y el posterior análisis del Assessment Experiences Questionnaire (AEQ) (Nuñez y Reyes, 2015). Sin embargo, éste no incluye dimensiones relevantes de los nuevos enfoques de enseñanza y aprendizaje, los cuales sitúan la evaluación al servicio del desarrollo del aprendizaje o de la adquisición de competencias (Carless, 2007).

Una de las dimensiones no incluidas en el AEQ es la de Agentes de Evaluación, que intenta recoger si el docente promueve el involucramiento de los estudiantes en la evaluación, incorporando autoevaluaciones y evaluación entre pares (Padilla Carmona y Gil Flores, 2008). Esta dimensión es importante, ya que los estudios señalan que los docentes deben diseñar y planificar evaluaciones, en las que juegue

un rol central la participación activa de los estudiantes en la evaluación, lo que dé cuenta de beneficios para el aprendizaje (Rodríguez-Gómez, Ibarra-Sáiz, Gallego-Noche, Gómez-Ruiz, y Quesada-Serra, 2012).

Otra dimensión que no se incluye en el AEQ y que se ha incorporado en el EEA, es Momentos de la Evaluación, esto es, la recogida de información en distintos momentos del proceso formativo, inicial, a lo largo del proceso o evaluación continua y final (Crisp, 2012), dimensión fundamental que hace posible una retroalimentación efectiva.

En definitiva, la escala EEA permite conocer el proceso de evaluación del aprendizaje de los estudiantes desde la experiencia de los propios estudiantes, la cual es relevante por su repercusión en el proceso y en los resultados de aprendizaje (Bordas y Cabrera, 2001; Padilla Carmona y Gil Flores, 2008; Van Gennip, Segers, y Tillema, 2009), lo que constituye una importante contribución, ya que existen pocas escalas de medida para conocer la evaluación de los aprendizajes en el aula (Pat-El et ál., 2013).

El presente estudio tuvo como objetivo principal diseñar y validar un instrumento que permitiera conocer el proceso de evaluación de los aprendizajes, desde un enfoque de Evaluación Orientado al Aprendizaje, que no sólo facilite el diagnóstico de las practicas evaluativas realizadas por los docentes, sino que también permita conocer el ajuste de las prácticas de evaluación a los cambios en los procesos curriculares, producto del tránsito de modelos de enseñanza basados en contenidos a modelos basados en competencias.

Por último, deben mencionarse algunas limitaciones del estudio. Por un lado, en relación con las dimensiones consideradas, “Agentes de Evaluación” solo incorpora dos ítems, por lo que debería reforzarse con más enunciados para mejorar la varianza explicada por la dimensión.

Por otro lado, la aplicación de la escala en una única Universidad supone que las características culturales de la muestra, el sistema de enseñanza y la formación docente, podrían limitar en cierta medida su validez en otros contextos; esto es, las posibilidades de generalización. Sin embargo, este estudio supone un avance significativo en el área de la evaluación con una aportación sustancial que mejoraría con sucesivos estudios en otras poblaciones. Así, la mejora futura de la escala debería incluir su adaptación y aplicación a otros contextos culturales y sistemas formativos.

## Referencias bibliográficas

- Álvarez Valdivia, I. (2008). Evaluación del aprendizaje en la universidad: Una mirada retrospectiva y prospectiva desde la divulgación científica. *Electronic Journal of Research in Educational Psychology*, 6(14), 235-271.
- Ballantyne, R., Hughes, K., y Mylonas, A. (2002). Developing procedures for implementing peer assessment in large classes using an action research process. *Assessment & Evaluation in Higher Education*, 27(5), 427-441.
- Barclay, D., Higgins, C., y Thompson, R. (1995). The partial least squares (PLS) approach to causal modeling: Personal computer adoption and use as an illustration. *Technology Studies*, 2(2), 285-309.
- Berliner, D.C. (1987). But do they understand? In V. Richardson-Koehler (Ed.), *Educator's Handbook: A Research Perspective* (pp. 259-293). New York: Longman
- Biggs, J. (2007). *Teaching for Quality Learning at University*. 2ª ed. New York: Open University Press.
- Bordas, M. I., y Cabrera, F. (2001). Estrategias de evaluación de los aprendizajes centrados en el proceso. *Revista Española de Pedagogía*, 218, 25-48.
- Boud, D., y Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31(4), 399-413.
- Broabfoot, P., y Black, P. (2004). Redefining assessment? the first ten years of Assessment in Education. *Assessment in Education*, 11(1), 1-27.
- Carless, D. R., Joughin, G. R., y Mok, M. M. C. (2006). Learning-oriented assessment: Principles and practice. *Assessment & Evaluation in Higher Education*, 31(4), 395-398.
- Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International*, 44(1), 57-66.
- Castro, M. C. B., Carrión, G. A. C., y Roldán, J. L. (2007). Investigar en economía de la empresa: ¿Partial least squares o modelos basados en la covarianza?. In *Proceedings El comportamiento de la empresa ante entornos dinámicos: XIX Congreso Anual y XV Congreso Hispano Francés de AEDEM* (p. 63). Asociación Española de Dirección y Economía de la Empresa (AEDEM).
- Contreras, G. (2010). Diagnóstico de dificultades de la evaluación del aprendizaje en la universidad: Un caso particular en Chile. *Educación y Educadores*, 13(2), 219-238.
- Chin, W. W. (2010). How to Write Up and Report PLS Analyses. In V. E. Vinzi, W. W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of Partial Least Squares* (pp. 655-690). Springer Berlin Heidelberg. Recuperado de [http://link.springer.com/chapter/10.1007/978-3-540-32827-8\\_29](http://link.springer.com/chapter/10.1007/978-3-540-32827-8_29)

- Churchill Jr., G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research (JMR)*, 16(1), 64-73.
- Crisp, G. T. (2012). Integrative assessment: Reframing assessment practice for current and future learning. *Assessment & Evaluation in Higher Education*, 37(1), 33-43. doi:10.1080/02602938.2010.494234.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334. doi:10.1007/bf02310555.
- DeVellis, R.F. (2003). *Scale development: theory and applications (2nd Ed.)*. Thousand Oaks: Sage Publications.
- Fernández, A. (2010). La evaluación orientada al aprendizaje en un modelo de formación por competencias en la educación universitaria. REDU. *Revista de Docencia Universitaria*, 8(1), 11.
- Fornell, C., y Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50. doi:10.2307/3151312
- Gefen, D., Straub, D. W., y Boudreau, M.-C. (2000). Structural equation modeling and regression: Guidelines for research practice. *Communications of the Association for Information System*, 4, 1-79.
- Gibbs, G., y Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education*, 1(1), 3-31.
- Gibbs, G., y Simpson, C. (2003). Measuring the response of students to assessment: The Assessment Experience Questionnaire. Paper presented at the proceedings of the 11th International Improving Student Learning Symposium, Hinckley, UK.
- Gil-Flores, J. (2012). La evaluación del aprendizaje en la universidad según la experiencia de los estudiantes. *Estudios sobre educación*, 22, 133-153.
- Hair, J. F. (1999). *Análisis multivariante*. Madrid, Spain: Prentice Hall.
- Hattie, J., y Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Hernandez, R. (2012). Does continuous assessment in higher education support student learning? *Higher Education: The International Journal of Higher Education and Educational Planning*, 64(4), 489-502.
- Howat, G., y Assaker, G. (2013). The hierarchical effects of perceived quality on perceived value, satisfaction, and loyalty: Empirical results from public, outdoor aquatic centres in Australia. *Sport Management Review*, 16(3), 268-284. doi:10.1016/j.smr.2012.10.001
- Ibarra, M<sup>a</sup> S. y Rodríguez, G. (2010). Aproximación al discurso dominante sobre la evaluación del aprendizaje en la universidad. *Revista de Educación*, 351, 385-407.

- Keppell, M., Au, E., Ma, A., y Chan, C. (2006). Peer learning and learning-oriented assessment in technology-enhanced environments. *Assessment & Evaluation in Higher Education*, 31(4), 453-464.
- Knight, P. (2006). The local practices of assessment. *Assessment & Evaluation in Higher Education*, 31(4), 435-452.
- Levy, J. P. L., Mallou, J. V., y González, J. A. (2003). *Análisis Multivariable para las Ciencias Sociales*. Madrid: Pearson Educación.
- Linn, R. L., Baker, E. L., y Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20(8), 15-21.
- Little, R.J.A. y Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons, Inc.
- Malhotra, N. K., Martínez, J. F. J. D., y Rosales, M. E. T. (2004). *Investigación de mercados: un enfoque práctico*. México: Pearson Educación.
- Manrique Villavicencio, I. (2009). La evaluación procesal y su rol en el cambio en la educación superior. *Educación* 18(34), 19-38.
- MacKenzie, S.B., Podsakoff, P.M., y Podsakoff, N.P. (2011). Construct measurement and validation procedures in MIS and behavior research: Integrating new and existing techniques. *MIS Quarterly*, 35, 293-334.
- Murphy, R. (2006). Evaluating new priorities for assessment in higher education. *Innovative Assessment in Higher Education*, 37-47.
- Navarro Soria, I. J., y González Gómez, C. (2010). La autoevaluación y la evaluación entre iguales como estrategia para el desarrollo de competencias profesionales. Una experiencia docente en el grado de maestro. REDU. *Revista de Docencia Universitaria*, 8(1), 187.
- Nunnally, J. (1978). *Psychometric theory (2nd ed.)*. New York, USA: McGraw-Hill.
- Núñez, J. L., y Reyes, C. I. (2015). La evaluación del aprendizaje de estudiantes: Validación española del assessment experience questionnaire (AEQ). *Estudios Sobre Educación*, 26, 63-77.
- Padilla Carmona, M. Teresa, y Gil Flores, Javier. (2008). La evaluación orientada al aprendizaje en la educación superior: Condiciones y estrategias para su aplicación en la docencia universitaria. *Revista Española de Pedagogía*, 241, 467-485.
- Pat-El, R. J., Tillema, H., Segers, M., y Vedder, P. (2013). Validation of assessment for learning questionnaires for teachers and students. *British Journal of Educational Psychology*, 83(1), 98-113.
- Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research*, 21(2), 381.

- Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly*, 31(4), 623-656.
- Popham, J. (2008). *Transformative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Rodríguez-Gómez, G., Ibarra-Sáiz, M. S., Gallego-Noche, B., Gómez-Ruiz, M., y Quesada-Serra, V. (2012). La voz del estudiante en la evaluación del aprendizaje: Un camino por recorrer en la universidad. *Relieve*, 18(2), 133-153.
- Ricoy, M. C. y Fernández-Rodríguez, J. (2013). La percepción que tienen los estudiantes universitarios sobre la evaluación: Un estudio de caso. *Educación XX1*, 16(2), 321-342. doi:10.5944/educxx1.16.2.2645.
- Sadler, D. R. (2010). Beyond feedback, developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35, 535-550.
- Santos Guerra, M.A. (1993). *La evaluación: un proceso de diálogo, comprensión y mejora*. Granada: Aljibe.
- Segura, M. (2007). La perspectiva ética de la evaluación de los aprendizajes desde un enfoque constructivista. *Actualidades Investigativas en Educación*, 7(1), 1-22.
- Sheikh, M. (2013). A comparative study of the assessment practices and proposed curriculum objectives in revised teacher education programs. *Bulletin of Education and Research*, 35(2), 75-89.
- Shepard, L. A. (2000). The Role of Assessment in a Learning Culture. *Teaching and Learning, Educational Research*, 29(7), 229-253.
- Tillema, H., Leenknecht, M., y Segers, M. (2011). Assessing assessment quality: Criteria for quality assurance in design of (peer) assessment for learning-a review of research studies. *Studies in Educational Evaluation*, 37(1), 25-34.
- Van Gennip, N. A. E., Segers, M. S. R., y Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review*, 4(1), 41-54.
- Villardón Gallego, M. L. (2006). Evaluación del aprendizaje para promover el desarrollo de competencias. *Educatio Siglo XXI*, 24, 57-76.
- William, D. (2011). What is assessment for learning? *Studies in Educational Evaluation* 37, 3-14.
- Wilson, M., y Scalise, K. (2006). Assessment to improve learning in higher education: The BEAR assessment system. *Higher Education*, 52(4), 635-663.