



**EVALUACIÓN
DEBATE 2014**



ESCALAS DE EVALUACIÓN Y AUTOEVALUACIÓN DE COMPETENCIAS DOCENTES EN LA UNIVERSIDAD DEL VALLE DE MÉXICO

Luis Pereyra Zetina
lpereyra@uvmnet.edu

Natalia Lima-Villeda
diana.limav@uvmnet.edu

Benjamín Armenta
benjamin.armenta.uvm@gmail.com

Resumen

El Sistema Integral de Evaluación del Desempeño Docente (SIEDD) de la Universidad del Valle de México (UVM) busca ser un proceso de evaluación sistemática. Con miras a la actualización de este sistema se diseñaron, validaron y confiabilizaron dos escalas de evaluación y dos de autoevaluación docente. A partir de esta experiencia, el objetivo del presente documento es describir la elaboración de dichas escalas. El procedimiento se dividió en cuatro fases con diferentes participantes y actividades en cada una de ellas. En la fase de análisis se aplicaron a los datos el coeficiente de alpha de cronbach ($\alpha > 0.8$, $\alpha > 0.9$) y el análisis factorial. Los resultados sugieren que las escalas pueden ser consideradas válidas y confiables.

Palabras clave: evaluación docente, escalas, elaboración de instrumentos

Planteamiento del problema

A partir de la actualización del SIEDD por medio de la construcción de escalas de evaluación, el planteamiento del problema en el presente estudio es ¿Cuál fue el proceso de construcción de escalas de evaluación de competencias docentes para el SIED en la UVM?

Objetivo

Describir el proceso de elaboración de dos escalas de evaluación y dos de autoevaluación de competencias docentes en los distintos niveles y modalidades educativos: licenciatura presencial (Ls) y licenciatura ejecutiva y posgrado (Lx-posgrado), para el SIED en la UVM.



EVALUACIÓN DEBATE 2014



Justificación

El Sistema Integral de Evaluación del Desempeño Docente (SIEDD) de la Universidad del Valle de México (UVM) evalúa sistemáticamente el desempeño de los docentes. El proceso se encamina a recabar información sobre las actividades académicas y administrativas en las que el profesor participa. Los propósitos son: proporcionar al docente retroalimentación sobre su desempeño, tomar decisiones tanto de la formación como actualización docente, impactar en el aprendizaje de los estudiantes y contribuir en la mejora de la calidad académica (Pereyra, 2014).

El SIEDD implementa métodos mixtos de evaluación, es decir, se emplean diversas fuentes, técnicas e instrumentos para el levantamiento de información. Con miras a perfeccionar el SIEDD y los instrumentos de evaluación del desempeño docente en la universidad, se han elaborado instrumentos que proporcionen información válida y confiable, para dar información que permita fortalecer al cuerpo docente de la institución, a través de programas de formación docente y actualización profesional (Pereyra, 2014).

En ese contexto se diseñaron, confiabilizaron y validaron dos escalas de evaluación y dos de autoevaluación docente. Dos de ellas corresponden a la evaluación docente por parte de estudiantes en las modalidades Ls y Lx-Posgrado; las otras dos corresponden a la autoevaluación docente en los mismos niveles y modalidades educativas.

En el presente estudio se muestra el proceso de elaboración y análisis psicométricos para su validación y confiabilización.

Fundamentación Teórica

Actualmente uno de los retos para las Instituciones de Educación Superior es ofrecer servicios de calidad. Uno de ellos está encaminado en brindar a los estudiantes espacios de desarrollo de competencias profesionales, necesarias para una inserción exitosa en el campo laboral. Siendo uno de los factores que coadyuvan en el logro de la calidad de los servicios, son sin lugar a dudas: las competencias docentes (Pereyra, 2014).



El término competencias docentes refiere al conjunto de conocimientos, habilidades y actitudes del docente, que se relacionan entre sí durante el proceso educativo de tal forma que, culminan con el desempeño exitoso tanto de las actividades como las funciones de la práctica docente. De tal forma que las capacidades docentes están relacionadas con tres aspectos: relaciones socioafectivas-motivacionales, proceso de enseñanza aprendizaje y con aspectos específicos de la disciplina que se enseña (Guzmán y Lima, 2013).

En la UVM se ha buscado promover el desarrollo de las competencias docentes, de tal forma, que, la evaluación del desempeño de los profesores es necesaria para contar con información que permita guiar los procesos de mejora pertinentes.

Evaluación docente

En el presente estudio se parte de la concepción de evaluación docente como la emisión de juicios de valor sobre la actividad docente en un periodo determinado, en el contexto de las metas y resultados esperados por la institución. La finalidad está encaminada al mejoramiento de la práctica docente (Pereyra, 2014).

Considerando que la evaluación docente es un proceso complejo que atiende diversos elementos, requiere de la integración de diversas formas y técnicas para recabar información. Las formas más utilizadas en evaluación del desempeño docente se pueden observar en la Tabla 1 (CENIDET 2005 - 2007).

Tabla 1
Formas utilizadas en la evaluación del desempeño docente

Formas de evaluación	Técnicas o instrumentos
Autoevaluación	Carpeta o portafolio, escalas
Evaluación de pares	Encuesta, entrevista, observación en el aula



Evaluación por los alumnos

Encuesta

Evaluación a través del
aprovechamiento escolar

Exámenes colegiados

Métodos mixtos

Empleo de diversas técnicas e
instrumentos

Nota. Recuperado de Pereyra (2014)

El SIEDD ha adoptado los métodos mixtos de evaluación docente, dado que ha sido la forma más adecuada de recabar información diversa y complementaria del desempeño de los docentes. Debido a que es un proceso que emplea distintas técnicas, instrumentos y fuentes de información, el presente documento se enfoca a describir la experiencia de elaborar escalas para este fin.

Escalas de evaluación

El objetivo de cualquier escala es medir alguna característica de las personas, es decir, asignar numerales o cuantificar la magnitud con que una persona presenta algún rasgo en particular. Sin embargo, dicha asignación no puede ser arbitraria, debe seguir un proceso que permita obtener como resultado un isomorfismo entre los números asignados y la magnitud del rasgo observado, de tal manera que, la covariación entre el resultado de las mediciones y el grado en que se presenta aquello que se mide sirva como un indicador fiable (Reidl, Guillén, Sierra y Joya, 2002).

Tal y como Reidl, Guillén, Sierra y Joya (2002) describen, el inicio de la construcción de una escala puede ser entendida mediante cuatro grandes fases: 1) la revisión de la literatura referente o relacionada al rasgo que se desea medir; 2) la identificación de las distintas dimensiones que pueden fungir como componentes de aquello para lo que se construye la escala; 3) la identificación y selección de los distintos indicadores para cada una de las distintas dimensiones; 4) la síntesis de la información con que se cuenta.

Después de la construcción de un marco teórico sólido, es posible elaborar los distintos reactivos que compondrán las dimensiones identificadas. La pertinencia de los reactivos, su lógica, su ordenamiento,



una clara redacción y un sistema de respuesta adecuado, son características fundamentales para lograr que la escala cumpla su objetivo (Cohen y Swerdlik, 2006).

Como todo instrumento de medición, una escala debe contar con dos grandes características: confiabilidad y validez. La confiabilidad refiere la consistencia o estabilidad de las mediciones, mientras que la validez es el grado con que un instrumento mide el rasgo que pretende medir (Cohen y Swerdlik, 2006). Una vez hecho un piloteo con una muestra de la población a quien va dirigido el instrumento, la confiabilidad puede ser evaluada de distintas maneras, como son: formas paralelas o alternativas, test-retest, división en dos partes y alpha de Cronbach, entre otras (Cohen y Swerdlik, 2006). Por ejemplo, si se asume que una escala confiable producirá mediciones similares, entonces se puede esperar que los resultados de dos mediciones hechas en distintos momentos con la misma muestra se correlacionen, a este método de confiabilización se le conoce como test-retest.

En la literatura, el método más comúnmente empleado es la obtención del Alpha de Cronbach, misma que constituye un indicador del grado con que se correlacionan los ítems que constituyen una escala (Cohen y Swerdlik, 2006).

Respecto a la validez, existen distintos tipos, mismos que cuentan con procedimientos específicos para obtenerlos. Los tres tipos de validez más importantes son: contenido, criterio y constructo (Cronbach y Meehl, 1955).

La validez de contenido hace referencia a la suficiencia y pertinencia de los elementos considerados como componentes representativos de una característica. Una manera de obtener este tipo de validez es mediante el juicio de expertos, personas especializadas en un ámbito de estudio relacionado estrechamente con el rasgo que se desea medir, los cuales evalúan tanto la pertinencia como la suficiencia de las dimensiones e ítems que constituyen una escala.

Por su parte, la validez de criterio es una comparación entre los puntajes de la variable de estudio con una variable criterio o externa. Se obtiene a partir de la correlación entre la medida obtenida con otras medidas



con las que teóricamente se correlacionaría, ya sea en sentido positivo o negativo (Prieto y Delgado, 2010).

Por último, la validez de constructo se llega a interpretar como un concepto unificador de la validez de contenido y de criterio, refiriendo en última instancia un análisis de la significación de los datos obtenidos a partir del instrumento de medición en relación a los conceptos asumidos (Pérez-Gil, Chacón y Moreno, 2000). Existen distintas formas de obtenerla, por ejemplo, mediante la identificación de diferencias entre poblaciones; es decir, si teóricamente una población debiera obtener un resultado menor o mayor a una población distinta, la constatación de dicha diferencia al aplicar el instrumento construido es una manera de obtener este tipo de validez.

Otro método usado con frecuencia para obtener validez de constructo es el análisis factorial, este puede servir para verificar o rechazar hipótesis acerca de la manera en que se agrupan los factores de los que se compone un rasgo (Cronbach y Meehl, 1955).

Es importante aclarar que contar con un tipo de validez, no constituye validez completa o suficiente. Por el contrario, la validación es un proceso continuo mediante el cual se recaban pruebas a favor del uso de una escala, por ello es conveniente contar con diferentes tipos de validez que arrojen información respecto a los cambios que son necesarios realizar.

Objetivo

Describir el proceso de elaboración de dos escalas de evaluación y dos de autoevaluación de competencias docentes en las distintos niveles y modalidades educativas: licenciatura presencial (Ls) y licenciatura ejecutiva y posgrado (Lx-posgrado), para el SIEDD de la UVM. Objetivo enmarcado dentro del proceso de construcción de escalas, conjugado con la noción de competencia así como el desempeño y la evaluación docente.

Método

Participantes



En la fase de validación de contenido los participantes fueron 3 expertos en investigación educativa y 59 docentes de distintos campus de la UVM. A su vez en la fase de piloteo participaron 667 estudiantes: 385 de Lx-posgrado y 282 de Ls. Así como 172 docentes: 59 de Lx-posgrado y 113 de Ls.

Procedimiento

La construcción de las escalas se realizaron en diferentes momentos: documentación y conceptualización, construcción, validación y ensayo de la prueba para análisis psicométricos. A continuación se describen los pasos de cada fase.

1. Documentación y conceptualización: Para la conceptualización de las categorías a medir en la escalas de evaluación y autoevaluación docente se revisaron documentos institucionales como el Modelo Educativo UVM, donde se prescribe el perfil del docente en la universidad, contrastando éste con trabajos de expertos en evaluación docente y estudios de otras instituciones educativas, con la finalidad de identificar los rasgos y atributos del docente en educación superior y media superior, así como las competencias del profesor y su comportamiento en el aula. A partir de esta revisión teórica, se definieron y conceptualizaron los elementos del SIEDD y las dimensiones que constituirían cada una de las escalas, a saber: a) Planeación del curso, b) Asiduidad, c) Gestión de los aprendizajes, d) Interacción en el aula e) Comunicación, e) Evaluación y f) Visión del curso.

2. Construcción: En la redacción de reactivos participaron tres funcionarios académicos cuya responsabilidad en la universidad es la gestión de los docentes con distintos niveles de responsabilidad y competencia organizacional. Para la realización de la tarea se les presentó la estructura y elementos del SIEDD, con la finalidad de contextualizarla y darle sentido. Asimismo, se les entregó una matriz con la nomenclatura y definición de las dimensiones a medir y el número de reactivos que deberían de contener éstas para cada una de las escalas. A partir del producto entregado por los redactores de reactivos, se construyó la primera versión del instrumento eligiendo, a juicio del responsable del proyecto, los reactivos mejor redactados y con mayor expresión de la categoría a medir.

3. Validación: El proceso de validación de contenidos se dio en primera instancia con tres jueces quienes evaluaron las instrucciones y cada reactivo en términos de claridad, pertinencia, redacción y relevancia. Después de la revisión de los expertos se realizaron las modificaciones necesarias. En un segundo momento, el proceso de validación de contenidos se realizó con 59 jueces, quienes evaluaron cada



reactivo en términos de claridad, pertinencia, redacción y relevancia. Después de la revisión de los docentes se realizaron las modificaciones correspondientes.

4. Ensayo de la prueba para análisis psicométricos: Después de la aplicación de las escalas a estudiantes y profesores se obtuvieron los análisis psicométricos de alpha de cronbach para obtener la confiabilidad por escala y subescalas, así como el análisis factorial para la validación de constructo.

Resultados

Validación de contenido

En la primera validación, los tres expertos en investigación educativa comentaron la necesidad de revisión de las instrucciones y consideraron adecuados a los 55% de los reactivos.

En la segunda validación con 59 docentes de diferentes campus de UVM, el 84.4% de los reactivos se consideran adecuados.

Análisis psicométricos

Cada una de las escalas fue sometida al análisis del coeficiente alpha de cronbach, obteniendo los resultados que se muestran en la tabla 2.

Tabla 2
Alpha de Cronbach por escala

Instrumento	α general
Autoevaluación LS	0.871
Autoevaluación Lx y posgrado	0.853
Evaluación LS	0.919
Evaluación Lx y posgrado	0.903

Nota: Las escalas cuentan con buen ($\alpha > 0.8$) o excelente nivel de confiabilidad ($\alpha > 0.9$)



Por el comportamiento de los ítems en el alpha (véase tabla 3), es recomendable eliminar o replantear las subescalas: Asiduidad, Satisfacción del curso y Evaluación.

Tabla 3
Alpha de Cronbach por subescala

Instrumento	Subescala (α)
Autoevaluación LS	Asiduidad (.394)
	Evaluación (.375)
Autoevaluación Lx y posgrado	Satisfacción del curso (-0.054)
Evaluación LS	Asiduidad (0.405)
Evaluación Lx y posgrado	Satisfacción del curso (-un solo ítem likert-)

Además, los datos de cada subescala fueron sometidos al análisis de factores (véase tabla 4), por el comportamiento de los ítems en el análisis factorial, es recomendable replantear las subescalas conforme se agrupan los ítems por factor

Tabla 4
Análisis factorial

Instrumento	Factor	% de varianza explicada posterior a la rotación	% acumulado de varianza explicada posterior a la rotación
Autoevaluación LS	1	15.768	51.772
	2	14.721	
	3	12.041	
	4	9.241	



EVALUACIÓN DEBATE 2014



Autoevaluación LX y Posgrado	1	27.981	56.438
	2	14.526	
	3	13.932	
Estudiantes LS	1	19.546	52.442
	2	19.308	
	3	13.588	
Evaluación LX y Posgrado	1	50.401	50.401

Nota: Número de factores obtenidos por escala y varianza explicada por cada uno de ellos.

Conclusiones

El objetivo del presente estudio fue describir el proceso de elaboración de dos escalas de evaluación y dos de autoevaluación de competencias docentes para las distintas modalidades y niveles educativos con los que cuenta la UVM.

El proceso de construcción siguió de forma óptima las etapas que suelen reportarse en la literatura, al contar con un marco teórico que diera sustento al contenido del instrumento, definiendo las escalas y subescalas para redactar los reactivos en cada una de ellas, así como la realización de la validez de contenido por jueces expertos. Para concluir con el cálculo del Alpha de Cronbach y el análisis factorial con el fin de obtener la confiabilidad y la validez de constructo.

Al obtener un Alpha de Cronbach por cada escala mayor a $\alpha 0.8$, se puede afirmar que la confiabilidad general de cada una de ellas fue buena o excelente, es decir, los instrumentos muestran consistencia interna.

Con respecto al análisis de factores obtenido, se recomienda agrupar los reactivos a partir de los resultados si se considera pertinente al objetivo del instrumento y su marco conceptual, ya que los factores para cada una de las escalas no correspondieron con los que se habían especificado a nivel teórico, en una primera fase.



En resumen, con la alta confiabilidad obtenida de todas las escalas y la validación por parte de los jueces expertos, se puede concluir que las escalas son adecuadas para generar información que permita tomar decisiones dentro de la institución.

En ese contexto es importante subrayar que todo proceso de evaluación de la práctica docente debe estar articulada a las finalidades institucionales, las cuales suponen definiciones claras del rol de cada uno de los agentes de la estructura institucional, además de la concepción clara del desempeño docente, con miras a que el proceso evaluativo conduzca a acciones de mejora (Pereyra, 2014).

Referencias bibliográficas

- Centro Nacional de Investigación y Desarrollo Tecnológico. (Julio, 2005). Evaluación del desempeño docente (CENIDET 2005-2007). Cuernavaca, Morelos. Recuperado de <http://www.cenidet.edu.mx/subaca/web-dda/docs/evaluaciondocente2005.pdf>
- Cohen, R. J. y Swerdlik, M. E. (2006). *Pruebas y Evaluación Psicológicas*. México: McGraw-Hill.
- Cronbach, L. J. y Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302. Recuperado de <http://psychclassics.yorku.ca/Cronbach/construct.htm>
- Guzmán-Cedillo, Y. I. y Lima, D. N. (2013) Los rankings universitarios y las competencias docentes: un análisis de las mejores universidades. Memorias VIII Congreso Internacional de Innovación Educativa. Recuperado en http://www.cgfie.ipn.mx/Paginas/Memorias_VIII_CIE.aspx
- Pereyra, L. (2014). Evaluación docente. Manuscrito inédito, Dirección de Evaluación Educativa, Universidad del Valle de México, Distrito Federal, México.
- Pérez-Gil, Chacón y Moreno, (2000). Validez de constructo: EL uso del análisis factorial exploratorio-confirmatorio para obtener evidencias de validez. *Psicothema*, 12(2), 442-446. Recuperado de <http://www.psicothema.com/pdf/601.pdf>
- Prieto, G. y Delgado, A. R. (2010). Fiabilidad y validez. *Papeles del Psicólogo*, 31(1), 67-74. Recuperado de <http://www.redalyc.org/pdf/778/77812441007.pdf>
- Reidl, L. M., Guillén, R., Sierra, G y Joya, L. (2002). *Celos y envidia: Medición alternativa*. México: UNAM.