

# Desarrollo de evidencias de validez de una prueba de desempeño sobre competencias docentes: EDOE

Facultad de Medicina



Marlette Lobato Valverde<sup>a</sup>, Amílcar Alpuche Hernández<sup>b</sup>,  
Juan Andrés Trejo Mejía<sup>b</sup>, Adrián Martínez González<sup>a,c,\*</sup>



## Resumen

**Introducción:** La evaluación de competencias docentes se relaciona con la mejora de la actividad educativa, orientada hacia el desarrollo personal y formativo. No se cuenta con evaluaciones de alta fidelidad en nuestro medio. El objetivo es obtener fuentes de evidencia de validez del Examen Docente Objetivo Estructurado (EDOE) para evaluar las competencias docentes.

**Método:** Se llevó a cabo un estudio cuantitativo en 3 fases: 1) Planeación: análisis del modelo de competencias, validación por jueces, validación de contenido por alumnos instructores y taller de trabajo con expertos para la creación de estaciones. 2) Piloto: se obtuvo el grado de concordancia de los evaluadores, validación y estructura interna de las estaciones (análisis de discriminación, alpha

de Cronbach, medias y desviación estándar). 3) Final: Aplicación a la muestra por conveniencia no aleatoria conformada por 20 profesores del Departamento de Informática Biomédica de la Facultad de Medicina de la UNAM.

**Resultados:** Se realizó un análisis de discriminación de reactivos de 8 estaciones 2 tienen 5 reactivos que discriminan .05, 7 obtienen cargas factoriales aceptables; explicando un 88.35% de la varianza, un alpha de Cronbach de .68. La media de desempeño docente en el EDOE fue de 73.23 con desviación estándar de 8.4. La estación mejor evaluada fue estrategias de aprendizaje ( $85 \pm 16.98$ ) y la menor fue orientar la discusión ( $67.48 \pm 15.62$ ). Se detectó evidencia confirmatoria con exámenes similares utilizados en la Facultad de Medicina,

<sup>a</sup>Departamento de Informática Biomédica, Facultad de Medicina, UNAM, Cd. Mx., México.

<sup>b</sup>Secretaría de Educación Médica, Facultad de Medicina, UNAM, Cd. Mx., México.

<sup>c</sup>Departamento de Salud Pública, Facultad de Medicina, UNAM, Cd. Mx., México.

Recibido: 30-octubre-2017. Aceptado: 22-enero-2018.

\*Autor para correspondencia: Adrián Martínez González. Ave. Universidad 3000. Circuito Escolar, C.U. Del. Coyoacán. Cd. Mx., México, 04510.

Correo electrónico: adrianmartinez38@gmail.com  
La revisión por pares es responsabilidad de la Universidad Nacional Autónoma de México.

2007-5057/© 2019 Universidad Nacional Autónoma de México, Facultad de Medicina. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).  
<http://dx.doi.org/10.22201/facmed.20075057e.2019.29.1779>

se cumplieron los objetivos planteados en la creación de este instrumento, se detectó una media preevaluación EDOE y los profesores fueron realimentados sobre su desempeño con fines formativos.

**Discusión y conclusiones:** Se propone un instrumento de evaluación EDOE con suficiente evidencia de validez que puede ser utilizado en la evaluación del grado de competencia docente. Lo anterior orienta a crear instrumentos de evaluación docente que sigan buenas prácticas en cuanto a la elaboración de instrumentos con base en la cadena de evidencia de validez.

**Palabras clave:** Examen Docente Objetivo Estructurado (EDOE); Validez; Competencia docente.

© 2019 Universidad Nacional Autónoma de México, Facultad de Medicina. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Development of validity evidences of a performance instrument on teaching competences: EDOE

### Abstract

**Introduction:** The evaluation of teaching competences is related to the improvement of the educational activity which is oriented towards personal and formative development. There are no high-fidelity evaluations in our specific setting. The aim was to obtain sources of evidence of validity for the Objective & Structured Teaching Examination (OSTE) to evaluate the teaching competencies.

**Method:** A quantitative study was carried out in three phases: 1) Planning phase: analysis of the competence model, validation by judges, validation of content by student instructors, and workshop with experts for the cre-

ation of the stations. 2) Pilot phase: to obtain the degree of concordance among evaluators, validation and internal structure of the stations (discriminant analysis, Cronbach's alpha, means and standard deviation), 3) Final phase: Application to a non-random sample formed by 20 professors of the Department of Biomedical Informatics at the Faculty of Medicine of UNAM.

**Results:** An item discrimination analysis of eight stations was carried out. We observed that two stations have five items that discriminate .05, seven had acceptable factorial loads and explained 88.35% of the variance with a Cronbach's alpha of .68. The average of teaching performance in the OSTE was 73.23 with a standard deviation of 8.4. The best-evaluated station was learning strategies ( $85 \pm 16.98$ ), and the lowest was discussion guidance ( $67.48 \pm 15.62$ ). Confirmatory evidence was obtained with similar examinations used in the Faculty of Medicine, the objectives set for the creation of this instrument were met, an OSTE pre-evaluation mean was detected, and the teachers got feedback on their performance for training purposes.

**Discussion and Conclusions:** The OSTE assessment is proposed as an instrument with sufficient evidence of validity that can be used in the evaluation of the degree of teaching competencies. We aim to create teaching evaluation instruments that follow good practices in the development of instruments based on the chain of evidence of validity.

**Keywords:** Objective Structured Teaching Exercise (OSTE); Validity; Teacher competences.

© 2019 Universidad Nacional Autónoma de México, Facultad de Medicina. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## INTRODUCCIÓN

La enseñanza de la medicina es una de las actividades en educación médica que ha tenido cambios vertiginosos e importantes en las últimas décadas<sup>1</sup>.

Esta actividad no solo implica el manejo de un solo proceso y de un apoyo constante disciplinal como fue en un inicio<sup>2-4</sup>. El repertorio de actividades de

enseñanza-aprendizaje y responsabilidades se transforman día a día<sup>5</sup>. En la actualidad hablamos de roles y competencias que un profesor desempeña al tener práctica docente<sup>6-8</sup>.

El docente requiere de competencias disciplinarias, psicopedagógicas, de comunicación, investigación, humanísticas y administrativas<sup>9</sup>. Estas deben

ser evaluadas con la finalidad de tomar decisiones sobre las áreas de mejora laboral y de su actividad educativa en el aula.

En este sentido, la evaluación es un proceso que intenta determinar sistemática y objetivamente la pertinencia y eficacia de la enseñanza-aprendizaje, y puede estar relacionada con la estructura, proceso, resultado y decisiones que se toman a través de ella<sup>10</sup>.

La evaluación de competencias docentes integra la evaluación en escenarios similares al real. En la literatura existen modelos referidos a la evaluación de las mismas. Además, existen pautas para la creación de instrumentos de evaluación del desempeño docente con evidencia de validez pero no de alta fidelidad en México.

La validez es un concepto ligado a la evaluación, consiste en el grado en el que la evidencia y la teoría apoyan las interpretaciones de los resultados de una prueba para los usos propuestos de una evaluación<sup>11</sup>. Se mencionan 5 diferentes fuentes de evidencia de validez<sup>12</sup> relacionadas con: el contenido del examen, el proceso de respuesta, la estructura interna, su relación con otras variables y las consecuencias para la persona que es objeto de la evaluación.

Hay pocos estudios sobre la construcción de instrumentos de alta fidelidad para el desempeño docente con pautas específicas de evidencia de validez. Uno de los modelos de alta fidelidad es el Examen Docente Objetivo Estructurado (EDOE)<sup>13</sup>.

Este fue descrito a principios de los años noventa al evaluar el desempeño docente de manera semejante a la competencia clínica, con base en el Examen Clínico Objetivo Estructurado (ECOE)<sup>14</sup>. Posteriormente Dunnington y DaRosa<sup>15</sup> aplicaron un examen con múltiples estaciones para evaluar a médicos residentes sobre diversos objetivos de un programa de “Residentes como maestros”. Además se propuso el uso de estudiantes estandarizados para evaluar el desempeño docente<sup>16,17</sup> y se realizaron revisiones sobre la validez y confiabilidad del EDOE concluyendo que se trata de un método de evaluación con validez y confiabilidad para evaluar competencias docentes<sup>18</sup>.

En la actualidad el EDOE es un modelo de evaluación integral, objetivo, válido, confiable y aceptable. Es un formato de evaluación que incorpora diversos instrumentos evaluativos y se desarrolla a lo largo de

varias estaciones que simulan situaciones contextualizadas de enseñanza estandarizada<sup>19</sup>. En el EDOE participan profesores que interactúan con estudiantes estandarizados y el evaluador. Permite valorar el logro de competencias específicas docentes, que se espera, debe poseer el profesor. Además, facilita evaluar las habilidades de enseñanza, mejora las destrezas docentes y valora la eficacia de los planes de estudio para potenciar las habilidades de enseñanza que se utilizan. Metodológicamente requiere de una planeación cuidadosa, diseño de estaciones, entrenamiento específico de los evaluadores y espacios físicos para efectuarlo<sup>20</sup>.

Se requieren más herramientas de evaluación sistemática, metódica e integral, además de los modelos en evaluación que existen actualmente. En escuelas de medicina mexicanas existe poca evidencia publicada del uso de este tipo de examen. El objetivo de este estudio es obtener fuentes de evidencia de validez de un instrumento para evaluar el desempeño docente.

La pregunta de investigación que guió el trabajo es ¿cuáles son las fuentes de evidencia de validez del Examen Docente Objetivo Estructurado (EDOE), para evaluar el grado de desempeño de los profesores de Informática Biomédica, de la Facultad de Medicina de la UNAM?

## MÉTODO

Se realizó un estudio no experimental de tipo descriptivo para la elaboración de los instrumentos de las estaciones y el proceso de evaluación a través del EDOE<sup>21,22</sup>. Se determinó la validez del EDOE en cada una de las fuentes de evidencia.

Se establecieron 3 fases: planeación, piloto y final, que tienen una secuencia metodológica y cronológica, para así poder maximizar la validez en cada una de sus fuentes.

Se utilizaron herramientas cualitativas para la fuente de validez de contenido. Para las fuentes de validez de proceso de respuesta y consecuencias se utilizaron herramientas cuantitativas. El contenido se basó en análisis de los escenarios educativos por parte de expertos en educación médica y alumnos que fueron reclutados con base en el criterio de tener al menos 3 años de experiencia en educación médica; el proceso de respuesta se integró con ca-

**Tabla 1.** Fases para el desarrollo de un Examen Docente Objetivo Estructurado

	<b>Fase</b>
<b>Preparación piloto final</b>	<b>Contenido:</b> Elementos del EDOE que representan el desempeño docente <ul style="list-style-type: none"> <li>• Construcción y representación del constructo <ul style="list-style-type: none"> <li>• Selección competencias docentes del Plan de Estudios 2010</li> <li>• Consulta a expertos con Metodología Delphi sobre las competencias para evaluar en el EDOE</li> <li>• Establecimiento de objetivos de medición (tipo y número de estaciones, duración, logística, tipo de instrumento de evaluación, tipo de evaluador, selección de actores)</li> </ul> </li> <li>• Taller de elaboración de material para estaciones del EDOE con expertos</li> <li>• Validación de contenido de las estaciones por parte de los alumnos</li> <li>• Entrenamiento y familiarización en el uso de listas de cotejo y guías de observación de los evaluadores en las tabletas IOS</li> <li>• Capacitación alumnos actores en las estaciones</li> <li>• Explicación de la logística de aplicación</li> </ul>
	<b>Proceso de respuesta:</b> Relación entre el constructo a medir y los participantes <ul style="list-style-type: none"> <li>• Realizar piloto del EDOE y registro de incidencias</li> <li>• Ajustar y modificar reactivos de los instrumentos de medición (listas de cotejo y guías de observación), estaciones y logística de acuerdo al registro de incidencias</li> <li>• Control de calidad en el vaciamiento de resultados</li> </ul>
	<b>Estructura interna:</b> Fiabilidad y estructura de factores <ul style="list-style-type: none"> <li>• Selección del análisis estadístico de acuerdo al rango de fiabilidad</li> </ul>
	<b>Relación con otras variables:</b> Búsqueda de evidencia confirmatoria <ul style="list-style-type: none"> <li>• Detección de pruebas similares en el medio organizacional</li> </ul>
	<b>Consecuencias de la prueba:</b> Evidencia para soportar interpretación y uso de los resultados para la toma de decisiones sobre el desempeño docente <ul style="list-style-type: none"> <li>• Revisión de los objetivos planteados</li> <li>• Establecimiento de umbral de puntuación</li> <li>• Determinar el uso de estos resultados: Evaluación formativa</li> </ul> Realimentación a los docentes con los resultados obtenidos a nivel individual

Tabla de construcción propia. Adaptada del artículo Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: Theory and application. Am J Med. 2006;119:165-6.

pacitaciones para familiarizarse con la forma de respuesta y la integridad de los datos tanto en papel como con el uso de un sistema informático; la estructura interna se definió por la valoración interjueces por parte los evaluadores aplicando  $\chi^2$  P de Pearson y la validación y estructura interna de las estaciones del examen (análisis de discriminación, alpha de Cronbach, medias y desviación estándar); para la relación entre variables se detectó evidencia confirmatoria con exámenes similares utilizados en la Facultad de Medicina. Por último, las consecuencias se definieron con estadística descriptiva del resultado global de la competencia docente como parte de la realimentación formativa a los profesores.

### Consideraciones éticas

Todos los participantes del estudio firmaron un consentimiento informado y tuvieron conocimiento de la finalidad de la investigación.

### ACTIVIDADES

Las actividades se desarrollaron en 3 fases de acuerdo a las 5 fuentes de evidencia de validez: a) planeación del EDOE, b) Prueba piloto y c) EDOE final. Para maximizar la validez del EDOE en cada una de sus fuentes, la preparación del EDOE incluyó diversas características (**tabla 1**).

#### a) Planeación del EDOE

Esta fase inició con un consenso de expertos donde se seleccionaron 49 de 79 actividades del Modelo

de Competencia Docente del profesor de medicina que sustenta las competencias docentes del Plan de Estudios 2010 de la Facultad de Medicina<sup>9</sup>. Posteriormente, a través de metodología Delphi se realizaron 2 rondas a través de Google Forms preguntando sobre las 49 actividades específicas por competencia. Participaron 11 expertos en cada una de ellas. La participación en ambas rondas fue del 100%. La primera ronda mostró porcentajes de acuerdo menores a 50% en algunos reactivos, por lo que se realizó una ronda más. En la segunda ronda el porcentaje aumentó al menos a 80% de acuerdo interjueces para establecer las actividades a considerar, con lo que se obtuvieron 12 competencias finales, de las cuales solo 8 eran observables y cuantificables para este tipo de evaluación<sup>23-25</sup>. Se realizó un taller sobre “Evaluación de la competencia docente con el Examen Docente Objetivo Estructurado” y una sesión de validación de contenido con alumnos instructores del Departamento de Informática Biomédica. Como resultado de estas sesiones se crearon 8 estaciones con los siguientes elementos:

- a) Libreto para los estudiantes.
- b) Instrumento de evaluación: Escala global para las estaciones dinámicas y lista de cotejo para las estáticas.
- c) Instrucciones para los profesores.
- d) Escenario de cada una de las competencias elegidas por los expertos (5 dinámicas y 3 estáticas), una duración de 8 minutos en cada una de ellas y que fueron denominadas:
  - 1) Manejo de fuentes información.
  - 2) Orientar la discusión.
  - 3) Evaluación.
  - 4) Estrategias educativas.
  - 5) Realimentar el aprendizaje.
  - 6) Estrategias de aprendizaje.
  - 7) Motivar a los alumnos.
  - 8) Conducir la dinámica del grupo.

Para atender la validez de la estructura interna, se crearon 2 tipos de instrumentos: guías de observación para las estaciones dinámicas y listas de cotejo para las estáticas. Las guías de observación inclu-

yerón de 5 a 7 atributos de la competencia docente a evaluar con la siguiente escala: A. Insuficiente, B. Suficiente, C. Bueno, D. Excelente y E. No Aplica. Siguiendo la sugerencia de Hodges et al.<sup>26</sup>, las listas de cotejo tenían 2 opciones (SÍ/NO) (**anexo**).

La metodología de diseño de estos documentos corresponde a una integración de las propuestas del ECOE<sup>27</sup>, EDOE<sup>17</sup>, de los pasos de NYU School of Medicine tanto para los formatos para la creación de las estaciones<sup>28</sup> como la integración del instrumento en general con 12 tips para el uso del EDOE de Bolliat<sup>29</sup>.

El proceso de respuesta se consideró de 3 maneras:

- 1) Capacitación del uso de los instrumentos de evaluación por parte de los evaluadores.
- 2) Tener 2 instructores por estación dinámica para evaluar a cada profesor.
- 3) Uso de tabletas Android para la captura de resultados.

Por último, sobre la validez asociada con las consecuencias de los sujetos evaluados, los profesores fueron informados de que el propósito de la prueba era formativo y de investigación educativa para la mejora continua del desempeño docente. Se contó con la ayuda del grupo de teatro de la Facultad de Medicina para la conformación de los escenarios de las estaciones.

### **b) Prueba piloto**

Se realizó una prueba piloto con el fin de mejorar la claridad y relevancia de las estaciones. Los participantes fueron 8 profesores de la Facultad de Medicina de la UNAM de diferentes departamentos. Se realizó en marzo del 2017 en el Departamento de Informática Biomédica. Se utilizaron hojas de lector óptico y de incidencias por estación. Posterior a la aplicación a través del análisis del registro de incidencias se realizaron modificaciones en las estaciones número 2 (orientar la discusión) y número 8 (conducir la dinámica del grupo) y en los instrumentos de las estaciones número 3 (evaluación educativa) y número 7 (motivar a los alumnos). La experiencia permitió que se ajustara el contenido de los casos y la actuación de los alumnos actores para la fase final.

**Tabla 2.** Concordancia por estación de la fase piloto entre evaluadores del Examen Docente Objetivo Estructurado

Estación	N	Evaluador 1	Evaluador 2	Correlación de Pearson
1	8	96.25	93.75	.96
2	8	63.33	69.16	.40
3	8	73.75	79.16	.45
5	8	65.00	54.16	.60
6	8	61.25	72.50	.72
7	8	77.08	67.50	-.34
8	8	56.14	29.47	.77

La tabla 2 muestra la concordancia por estación en la fase piloto entre ambos evaluadores del Examen Docente Objetivo Estructurado a través de la correlación de Pearson.

### c) EDOE final

La muestra estuvo conformada por 20 profesores del Departamento de Informática Biomédica de acuerdo a un muestreo por conveniencia. Los docentes representaron 30% de los profesores del Departamento de Informática Biomédica. La evaluación tuvo una duración 1 hora con 20 minutos. Cada estación tuvo una duración de 8 minutos. Se utilizaron tabletas Android cuya información fue capturada de manera instantánea al sistema Moodle. Posterior a la aplicación se realizó la extracción de los resultados de las tabletas y el registro y análisis del registro de incidencias. Las tabletas de la estación 4 no se registraron debido a una falla en el sistema quedando solo 7 estaciones para el análisis de resultados.

## RESULTADOS

Las características sociodemográficas de la muestra de 20 profesores del EDOE fueron las siguientes: 68% fueron hombres y 32% mujeres, la edad promedio fue de 39 años con una media de impartición de clases de 8 años.

### Planeación del EDOE

La validez de contenido se determinó por la elección de las competencias para cada una de las estaciones con metodología Delphi, el diseño de los casos y la prueba piloto para la preparación de documentos que dan sustento al contenido temático del examen y a la representatividad de dominios del área a evaluar y la calidad de las estaciones.

Del Modelo de Competencia Docente se seleccio-

naron 12 competencias, de acuerdo con su pertinencia, relevancia y claridad. De estas, se eligieron 8 por consenso de expertos, las cuales son observables y cuantificables en este tipo de evaluación.

### Piloto del EDOE

El proceso de respuesta resultó del proceso de validación de las estaciones, las guías de observación y las listas de cotejo. El instrumento resultado fue un Examen Docente Objetivo Estructurado conformado de 8 estaciones de las cuales 5 fueron dinámicas y 3 estáticas con guías de observación y listas de cotejo. Para el control de errores e integridad de los datos la administración de la prueba se realizó al 100% por los evaluadores a través de tabletas Android.

Para la estructura interna como parte de la estadística descriptiva, indica que la media de la competencia docente fue de 73.0 en escala de 0 a 100 con una desviación estándar de 8.4.

Se realizó un análisis de discriminación de reactivos de las 8 estaciones con sus respectivas escalas, se observa que de las 8 estaciones, 2 de estas tienen 5 reactivos que discriminan .05. Las restantes no presentan datos concluyentes en el análisis. Con respecto a la estructura de la prueba de las 8 estaciones propuestas solo 7 obtuvieron cargas factoriales aceptables, explicando un 88.3% de la varianza. Finalmente para la consistencia interna de la prueba se obtuvo un alpha de Cronbach de .68

Además se calculó para cada estación el índice de acuerdo entre evaluadores con la prueba  $\chi^2 P$  de Pearson por reactivo y estación.

**Tabla 3.** Concordancia por estación de la fase final entre evaluadores del Examen Docente Objetivo Estructurado

Estación	Final N	Eva 1	Eva 2	Correlación de Pearson
1	—	—	—	—
2	17	57.64	77.58	.69
3	19	84.73	80.00	.57
5	17	49.21	69.21	.41
6	—	—	—	—
7	20	84.66	66.66	.40
8	20	81.00	76.50	-.15

\*La estación 4 no está en el análisis.

La tabla 3 muestra la concordancia por estación en la fase final entre ambos evaluadores del Examen Docente Objetivo Estructurado.

**Tabla 4.** Resultados obtenidos por los profesores en el Examen Docente Objetivo Estructurado

Estación	Final N	Eva 1	Eva 2	Correlación de Pearson
1 Manejo de fuentes información	27	75.18	67.37	82.99
2 Orientar la discusión	28	67.48	61.42	73.53
3 Evaluación	28	81.30	73.95	88.66
5 Realimentar el aprendizaje	28	60.47	53.28	67.66
6 Estrategias de aprendizaje	27	85.00	78.28	91.71
7 Motivar a los alumnos	28	74.70	69.32	80.08
8 Conducir la dinámica del grupo	28	68.48	60.49	76.46
Total	28	73.23	63.30	80.15

\*Del intervalo de confianza al de 95%.

La tabla 4 muestra los resultados obtenidos por los profesores tanto del piloto y final del Examen Docente Objetivo Estructurado por estación y sus límites inferior y superior. La estaciones con mayor puntuación fueron: Estrategias de aprendizaje, evaluación, manejo de fuentes de información y motivar a los alumnos. Las estaciones con menor puntaje fueron realimentar, conducir la dinámica de grupo y orientar la discusión.

La **tabla 2** muestra la concordancia por estación en la fase piloto entre ambos evaluadores del Examen Docente Objetivo Estructurado a través de la correlación de Pearson. En general las puntuaciones van de moderadas a altas. A través de esta etapa se realizaron las correcciones a cada una de las estaciones en cuanto al contenido, el guión de los alumnos actores y los instrumentos de evaluación.

### Final del EDOE

La **tabla 3** muestra la concordancia por estación en la fase final entre ambos evaluadores del Examen Docente Objetivo Estructurado. En general las puntuaciones van de moderadas a altas. Las estaciones que presentan mayor número de acuerdo en las observaciones realizadas por los evaluadores son las

estáticas (2 y 3) como son: manejo de fuentes de información y evaluación. En tanto que la estación 8 de motivación tiene menor acuerdo.

La **tabla 4** muestra los resultados obtenidos por los profesores tanto del piloto como de la fase final del Examen Docente Objetivo Estructurado por estación y sus límites inferior y superior. La estaciones con mayor puntuación fueron: Estrategias de aprendizaje, evaluación, manejo de fuentes de información y motivar a los alumnos. Las estaciones con menor puntaje fueron realimentar, conducir la dinámica de grupo y orientar la discusión.

Para la relación con otras variables se encontró evidencia confirmatoria con exámenes similares en otras publicaciones. La media pre EDOE fue similar a la de otros estudios<sup>30,31</sup>.

Las consecuencias de la prueba se dieron a conocer con la entrega de los resultados al 100% de los profesores que participaron en el EDOE, realimentando así su desempeño exclusivamente para fines formativos.

## DISCUSIÓN

Por primera vez en México se evaluó la competencia docente con un instrumento de alta fidelidad con fuentes de evidencia de validez suficientes para la aplicación de un Examen Docente Objetivo Estructurado. Las fuentes de evidencia de validez corresponden a las mencionadas en los estándares de la American Educational Research Association, Downing, los parámetros de NYU SOM en cuanto a los formatos para la creación de las estaciones y los 12 *tips* para el uso del EDOE de Bolliat. El estudio aporta evidencia de validez en relación al contenido, proceso de respuesta, estructura interna, relación con otras variables y consecuencias.

La representatividad del contenido en comparación con otros estudios<sup>32-34</sup> tomó en cuenta diversas fuentes tanto de expertos como estudiantes de medicina con metodología cualitativa como la metodología Delphi.

Para el proceso de respuesta se realizó, como en otros estudios<sup>17,35-39</sup>, el entrenamiento para el uso de los instrumentos y del sistema informático para que la recolección fuera inmediata.

En cuanto a la estructura interna, diversos estudios examinan la fiabilidad de los componentes específicos del EDOE como medio para evaluar el rendimiento de la enseñanza, incluido el instrumento de calificación y el tipo de evaluador utilizado<sup>15,16,18,30</sup>. La mayoría de los trabajos cuantitativos que se han centrado en determinar la fiabilidad y validez del EDOE sugieren que es un instrumento fiable para evaluar las habilidades de enseñanza<sup>19</sup>. Las confiabilidades que se han reportado en los estudios con EDOE van de .63 a .98. Con correlaciones intraclase de .75 para cada una de las estaciones con el uso de varios observadores y con un descenso de la confiabilidad (0.61) al ser solo un evaluador<sup>40,41</sup>. Asimismo, se nota un descenso (.88) en la confiabilidad cuando se analizan las estaciones por separado<sup>42</sup>. En nuestro estudio se reporta un alpha de Cronbach de 0.68, con una media de desempeño docente de 73.2 semejante a lo que se ha reportado en la literatura.

Se han hecho estudios de seguimiento de un EDOE mostrando resultados comparables (alfa de Cronbach 0.82-0.90 para cada estación y 0.79 para toda la prueba)<sup>31</sup>. Asimismo, se han reportado instrumentos menos fiables<sup>15,16,43,44</sup>. La fiabilidad no varió en función del tipo de evaluador utilizado<sup>16,45</sup>. Otro estudio demostró que la fiabilidad del evaluador fue influenciada más por la formación en el EDOE que por la experiencia del evaluador<sup>44</sup>. Los estudiantes estandarizados eran más confiables que los evaluadores entrenados cuando se utilizó un medio de evaluación basado en la lista de verificación<sup>45</sup>. En nuestro caso utilizamos 2 tipos de instrumentos de evaluación, guías de observación y listas de cotejo con alumnos entrenados en el uso de estos instrumentos.

Para la relación con otras variables se encontró evidencia confirmatoria como medias y resultados con exámenes similares en otras publicaciones<sup>46,47</sup>. Es importante señalar que el uso del EDOE en nuestro medio no tiene aún ninguna aplicación para el desarrollo de un programa integral de evaluación docente.

Por último, las consecuencias de la prueba se dieron a conocer con la entrega de los resultados, realimentando así su desempeño exclusivamente para fines formativos<sup>48,49</sup>.

En términos generales, con respecto a la confiabilidad hay concordancia con resultados de la creación de estaciones para evaluar la competencia docente<sup>30,31</sup>. En cuanto a los estudios de validez realizados con el EDOE, estos han sido enfocados de manera individual en la validez de contenido en específico en relación con la opinión de los participantes sobre la aceptabilidad, el realismo de los escenarios de la enseñanza clínica con el EDOE<sup>14,16,36,50</sup>. Todos estos estudios señalaron que los participantes encontraron las situaciones EDOE realistas y sirvieron como una aproximación razonable de un encuentro de enseñanza. En una investigación de validez de contenido se sugirió el uso de grupos focales y se encontró un 92% de acuerdo entre los participantes<sup>18</sup>. La validez predictiva se evaluó a través de la capacidad del instrumento para detectar mejoras en el desempeño docente en comparación con los no expuestos.

Las fortalezas más importantes de nuestro estu-

dio son la obtención de fuentes de evidencia validez suficiente para la aplicación del EDOE en nuestro medio y la evaluación de competencias docentes con un instrumento de alta fidelidad que evalúa el constructo sobre el desempeño docente con la participación tanto de expertos como de alumnos de la facultad; relaciona el constructo a medir con los participantes tanto en las estaciones como en los usuarios de los instrumentos; selecciona y realiza un análisis de acuerdo al rango de fiabilidad; logra la evaluación de competencias docentes y realimenta de manera individual a los profesores que participaron en el estudio.

De acuerdo a los puntajes en los índices de concordancia con  $\chi^2$  de Pearson estos son mejorables: el proceso de respuesta entre alumnos evaluadores y la relación entre la percepción de profesores y actores. Esto indica que se requiere reforzar la capacitación y entrenamiento de los evaluadores alumnos para evaluar las actividades de cada una de las competencias<sup>51</sup> con el propósito de fortalecer la fuente sobre el proceso de respuesta, ya que es indispensable para fortalecer el objetivo particular de medición y ejemplos de cómo realizar un estudio de validación para obtener tales evidencias de validez.

Una de las limitaciones de nuestro estudio es que faltó evaluar la satisfacción de los profesores en su experiencia con el EDOE y la estrategia de realimentación por estación por parte de los evaluadores. Existen nuevas competencias que se están evaluando con el EDOE<sup>52</sup>, por lo que un nuevo estudio podría abarcar estas deficiencias.

Es cuanto a la estructura interna, nuestra alpha se encuentra dentro de los rangos inferiores de las que se han reportado en la literatura. Por lo que es necesario seguir realizando estudios en esta línea para evaluar de manera efectiva la dimensionalidad, invarianza de la medida y fiabilidad de un *test* educativo u otro tipo de evaluación como es nuestro caso<sup>51,53</sup>.

Esta línea de investigación marca el inicio de la creación de instrumentos de alta fidelidad en la evaluación del desempeño docente en ciencias de la salud.

## CONCLUSIONES

Los resultados descritos en el presente artículo permitieron evaluar la competencia docente con un

instrumento de alta fidelidad, de acuerdo a las recomendaciones que se proponen en la literatura, obteniendo evidencias de validez y confiabilidad relacionadas con el contenido del instrumento, el proceso de respuesta, estructura interna, relación con otras variables y consecuencias. Es importante realizar este tipo de instrumentos para la formación y evaluación de competencias en áreas de trabajo de formación y desarrollo docente.

## CONTRIBUCIÓN INDIVIDUAL

- MLV: Revisión de la información y articulación del manuscrito, operatividad de la aplicación del instrumento, análisis estadístico, correcciones y traducción en inglés.
- AAH: Revisión de la información, articulación del manuscrito y apoyo en análisis estadístico.
- JATM: Revisión de la información y articulación del manuscrito.
- AMG: Concibió y diseño el proyecto. Revisión de la información y articulación del manuscrito.

## AGRADECIMIENTOS

Al Departamento de Informática Biomédica y al grupo de teatro de la Facultad de Medicina.

## PRESENTACIONES PREVIAS

Ninguna.

## FINANCIAMIENTO

Ninguno.

## CONFLICTO DE INTERESES

Ninguno. 

## REFERENCIAS

1. Centra JA. Types of faculty development programs, J of Higher Ed. 1978;(49):151-62.
2. Benor DE. Faculty development, teacher training and teacher accreditation in medical education: twenty years from now, Med Teach. 2000;22:503-12.
3. Guskey TR. What makes professional development effective? Phi Delta Kappan. 2003;84:748-50.
4. Steinert Y. Staff development. En Harden R, editor A Practical Guide for Medical Teacher. 2005;390-9.
5. Harris DL, Krause KC, Parish DC, Smith MU. Academic competencies for medical faculty. Fam Med. 2007;39:343-50.
6. Spencer JA. Learner centred approaches in medical education. BMJ. 1999;318:1280-3.

7. Harden RM, Crosby JR. AMEE Education Guide No 20: The good teacher is more than a lecturer – the twelve roles of the teacher. *Med Teach.* 2000;22:334-47.
8. Bland CJ, Schmitz C, Stritter F, Henry R, Aluise J. Successful faculty in academic medicine: Essential skills and how to acquire them. New York, NY: Springer Publishing; 1990.
9. Martínez-González A, López-Bárcena J, Herrera Saint-leu P, et al. Modelo de competencias del profesor de medicina. *Educ Med.* 2008;11:157-67.
10. Linn RL, Gronlund NE. Measurement and Assessment in Teaching. Prentice-Hall, 8th Ed. 2000.
11. Haladyna TM, Downing SM, Rodriguez MC. A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Applied Measurement in Education.* 2002; 15:309-34.
12. Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ.* 2003;37:830-7.
13. Downing SM, Yudkowsky R. Introduction to Assessment in the Health Professions. En: Downing SM, Yudkowsky (Eds) *Assess in Health Prof Educ.* New York, NY: Routledge; 2009. p. 217-40.
14. Simpson D, Lawrence S, Krogull S. Using standardized ambulatory teaching situations for faculty development. *Teach Learn Med.* 1992;4:58-61.
15. Dunnington GL, DaRosa DA. A prospective randomized trial of residents as teachers training program. *Academic Medicine.* 1998;73,(6):696-700.
16. Prislin MD, Fitzpatrick C, Giglio M, Lie D, Radecki S. Initial experience with a multi-station objective structured teaching skills evaluation. *Acad Med.* 1998;73:1116-8.
17. Zabar S, Hanley K, Stevens D, Kalet A, Schwartz M, Pearlman E, Brenner, Kachur E, Lipkin M. Measuring the Competence of Residents as Teachers. *JGIM.* 2004;19:530.
18. Morrison EH, Boker JR, Hollingshead J, Prislin MD, Hitchcock MA, Litzelman DK. Reliability and validity of an objective structured teaching examination for generalist resident teachers. *Acad Med.* 2002;77:S29-S32.
19. Trowbridge RL, Snydman LK, Skolfield J, Hafler J, Bing-You RG. A systematic review of the use and effectiveness of the Objective Structured Teaching Encounter. *Med Teach.* 2011;33:893-903.
20. Sturpe DA et al. A Primer for Objective Structured Teaching Exercises. *American J of Pharma Educ.* 2014;78:1-6.
21. Creswell J. Choosing a mixed methods design. En: Designing and conducting. Mixed methods research. United States of America: SAGE. 2011:53-106.
22. Kerlinger FN, y Lee HB. Investigación del comportamiento. México: McGraw-Hill; 2002.
23. Thangaratnam S, Redman C. The Delphi technique. *The Obstetrician & Gynaecologist.* 2005;7:120-5.
24. Von der Gracht H. The future of logistics. Glaber Edition. *The Obstetrician & Gynaecologist.* 2008;120-5.
25. Varela-Ruiz et al. Descripción y usos del método Delphi en investigaciones del área de la salud. *Inv Ed Med.* 2012; 1(2):90-5.
26. Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise. *Acad Med.* 1999;74:1129-34.
27. Trejo-Mejía JA, Blee-Sánchez GB, Peña-Balderas J. Elaboración de estaciones para el examen clínico objetivo estructurado (ECOE). *RIEM.* 2013;3:56-9.
28. McAndrew M, et al. Creating an Objective Structured Teaching Examination to Evaluate a Dental Faculty Development Program. *J of Dental Educ.* 2012;76:461-8.
29. Boillat M. Twelve tips for using the Objective Structured Teaching Exercise for faculty development. *Med Teach.* 2012;34:269-73.
30. Schol S. A multiple-station test of the teaching skills of general practice preceptors in Flanders, Belgium. *Acad Med.* 2001;76:176-80.
31. Schol S, Goedhuys J, Notten T, Betz W. Individualised training to improve teaching competence of general practitioner trainers: A randomised controlled trial. *Med Educ.* 2005;39:991-8.
32. Maureen M, et al. Creating an Objective Structured Teaching Examination to Evaluate a Dental Faculty Development Program. *JDE.* 2012;76:461-8.
33. Deborah A, et al. An objective structured teaching examination (OSTE) for faculty development. *Curr in Phar Teach and Learn.* 2013;5:616-9.
34. Kachur EK. Multiple station exams and teaching exercises (MSEs) for teaching about teaching. In: Scherpel AJJA, van der Vleuten CPM, Rethans JJ, van der Steeg AFW (Eds). *Adv Med Educ.* Dordrecht: Kluwer Academic Publishers.1997. p. 226-7
35. Ellen J, Giardino AP, Edinburgh K, Ende J. Simulated students: A new method for studying clinical precepting. *Teach Learn Med.* 1994;6:132-5.
36. Gelula MH. Using standardized medical students to improve junior faculty teaching. *Acad Med.* 1998;73:611-2.
37. Gelula MH, Yudkowsky R. Microteaching and standardized students support faculty development for clinical teaching. *Acad Med.* 2002;77:941.
38. Lang F, Bennard B, Belanger A. Using standardized students to educate preceptors. *Acad Med.* 1995;70:855-6.
39. Pangaro LN, Worth-Dickstein H, Macmillan MK, Klass DJ, Shatz JH. Performance of “standardized examinees” in a standardized-patient examination of clinical skills. *Acad Med.* 1997;72:1008-11.
40. Morrison EH, Lewis EM, Gabbert CC, Boker JR, Kumar B, Harthill M. Evaluating a ‘service elective’ in clinical teaching for medical students. *Med Teach.* 2003;25:662-3.
41. Morrison EH, Rucker L, Boker JR, Gabbert CC, Hubbell FA, Hitchcock MA ,Prislin MD. The effect of a 13-hour curriculum to improve residents’ teaching skills: A randomized trial. *Ann Intern Med.* 2004;141:257-63.
42. Morrison EH, Rucker L, Boker JR, Hollingshead J, Hitchcock MA, Prislin MD, Hubbell FA. A pilot randomized, controlled trial of a longitudinal residents-as-teachers curriculum. *Acad Med.* 2003;78:722-9.

43. Stone S, Mazor K, Devaney-O'Neil S, Starr S, Ferguson W, Wellman S, Jacobson E, Hatem DS, Quirk M. Development and implementation of an objective structured teaching exercise (OSTE) to evaluate improvement in feedback skills following a faculty development workshop. *Teach Learn Med.* 2003;15:7-13.
44. Srinivasan M, Litzelman D, Seshadri R, Lane K, Zhou W, Bogdewic S, Gaffney M, Galvin M, Mitchell G, Treadwell P, et al. Developing an OSTE to address lapses in learners' professional behavior and an instrument to code educators' responses. *Acad Med.* 2004;79:888-96.
45. Quirk M, Mazor K, Haley HL, Wellman S, Keller D, Hatem D, Keller LA. Reliability and validity of checklists and global ratings by standardized students, trained raters, and faculty raters in an objective structured teaching environment. *Teach Learn Med.* 2005;17:202-9.
46. Wamsley MA, Julian KA, Vener MH, Morrison EH. Using an objective structured teaching evaluation for faculty development. *Med Educ.* 2005;39:1160-1.
47. Julian K, et al. The Impact of an Objective Structured Teaching Evaluation on Faculty Teaching Skills. *2012;24:3-7.*
48. Wilkerson L, Irby DM. Strategies for improving teaching practices: A comprehensive approach to faculty development. *Acad Med.* 1998;73:387-96.
49. Wilkes MS, Hoffman JR, Usatine R, Baillie S. An innovative program to augment community preceptors' practice and teaching skills. *Acad Med.* 2006;81:32-341.
50. Gelula MH, Yudkowsky R. Using standardised students in faculty development workshops to improve clinical teaching skills. *Med Educ.* 2003;37:621-9.
51. Padilla JL, Benítez I. Validity evidence based on response processes. *Psico.* 2014;26:136-44.
52. Cerronea S, et al. Using Objective Structured Teaching Encounters (OSTEs) to prepare chief residents to be emotionally intelligent leaders. *Med Educ online.* 2017;22:1-7.
53. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: Theory and application. *Am J Med.* 2006;119:165-66.

## Anexos

### Estación dinámica EDOE Conducir la dinámica del grupo INSTRUCCIONES PARA EL PROFESOR

Se encuentra en el aula con un grupo con el que lleva trabajando dos meses. Enseñará una habilidad práctica y requiere organizar al grupo en equipos. Sabe que su grupo no tiene una buena convivencia entre ellos.

A continuación, organice la actividad para que trabajen en equipos.

<b>GUÍA SINTÉTICA PARA EL EVALUADOR</b> <b>Conducir la dinámica del grupo</b>				
A continuación se presentan los dominios de la competencia docente, cada uno de ellos tiene los criterios a considerar para emitir un juicio con base en el desempeño del profesor				
Marque la opción elegida con una "X"			Ejemplo:	C. Bueno 
<b>1. Ambiente</b>				
<b>A. Insuficiente</b>	<b>B. Suficiente</b>	<b>C. Bueno</b>	<b>D. Excelente</b>	<b>E. No aplica</b>
NO crea un ambiente en el que los estudiantes pueden crecer personal y profesionalmente			Crea un ambiente en el que los estudiantes pueden crecer personal y profesionalmente	
<b>2. Desarrollo</b>				
<b>A. Insuficiente</b>	<b>B. Suficiente</b>	<b>C. Bueno</b>	<b>D. Excelente</b>	<b>E. No aplica</b>
NO proporciona oportunidades de desarrollo para replantear las metas individuales/grupales			Proporciona oportunidades de desarrollo para replantear las metas individuales/grupales	
<b>3. Identifica</b>				
<b>A. Insuficiente</b>	<b>B. Suficiente</b>	<b>C. Bueno</b>	<b>D. Excelente</b>	<b>E. No aplica</b>
NO crea o identifica oportunidades de desarrollo adaptadas a las metas profesionales			Crea o identifica oportunidades de desarrollo adaptadas a las metas profesionales	
<b>4. Perspectiva</b>				
<b>A. Insuficiente</b>	<b>B. Suficiente</b>	<b>C. Bueno</b>	<b>D. Excelente</b>	<b>E. No aplica</b>
NO amplia las perspectivas de los individuos o grupos			Amplia las perspectivas de los individuos o grupos	
<b>5. Realimentación</b>				
<b>A. Insuficiente</b>	<b>B. Suficiente</b>	<b>C. Bueno</b>	<b>D. Excelente</b>	<b>E. No aplica</b>
NO provee realimentación a estudiantes o grupos			Provee realimentación a estudiantes o grupos	
<b>6. Apoyo</b>				
<b>A. Insuficiente</b>	<b>B. Suficiente</b>	<b>C. Bueno</b>	<b>D. Excelente</b>	<b>E. No aplica</b>
NO identifica o desarrolla recursos para que los estudiantes alcancen sus metas			Identifica o desarrolla recursos para que los estudiantes alcancen sus metas	

## Estación estática EDOE

## Instrumentar las estrategias de aprendizaje adecuadas a la situación educativa.

## INSTRUCCIONES PARA EL PROFESOR

Se encuentra a mitad del semestre y sus alumnos no asistieron a clase porque presentaron departamental en Tlatelolco. A continuación escriba las estrategias de aprendizaje que usted realizaría para reponer la clase en línea.

### Lista de cotejo para el evaluador

## Lista de cotejo para el evaluador

Instrucciones: Marque la opción elegida con una "X".		Ejemplo:		No
Enunciado		Sí	No	
Planteó una estrategia de aprendizaje				
Describió la estrategia de aprendizaje adecuada a la situación				
Describió el uso del aula virtual o medios digitales				
Utilizó recursos en línea				
Utilizó elementos pedagógicos de diseño de clase				