



XVI
Congreso Nacional de
Investigación Educativa
CNIE-2021

Tres décadas de investigación de la evaluación de la docencia: algunas perspectivas

Mario Rueda Beltrán
IISUE-UNAM
mariorb@unam.mx

Balance y perspectivas sobre la evaluación de la docencia

Alejandro Canales
IISUE-UNAM
canalesa@unam.mx

Experiencia de metaevaluación de un sistema de evaluación docente. Retos y perspectivas

Edna Luna Serrano
Universidad Autónoma de Baja California, México
eluna@uabc.edu.mx

El contexto en la evaluación de competencias profesionales: avances conceptuales y metodológicos

Dra. Yolanda Edith Leyva Barajas
Centro de estudios sobre la Enseñanza y el Aprendizaje del Derecho
yolandalb@ceead.org.mx



Área temática 12. Evaluación educativa.
Línea temática: Evaluación docente y desarrollo profesional.

Resumen general del simposio

Una característica de la investigación educativa, como actividad profesional reconocida en nuestro país, ha sido el estudio de los temas que por alguna razón han sido significativos para el sistema educativo. Tal es el caso, por ejemplo, de la evaluación, siempre presente en el campo educativo, pero que, sin embargo, tuvo un crecimiento notable cuando formó parte nodal de las políticas gubernamentales. Si el tema de la evaluación educativa formaba parte de los trabajos sobre planeación, ahora es tal la cantidad de investigaciones sobre el tema que el COMIE ha creado un área especial para ubicar toda la producción actual. También la evaluación de la docencia, con una presencia discreta hace tres décadas, a partir de la reforma que la ubicó como condición de ingreso y permanencia del profesorado de educación básica, volcó el interés de un mayor número de académicos en estudiar de manera sistemática el impacto que estas políticas han tenido en el quehacer profesional de los trabajadores de la educación.

La investigación educativa, puede contribuir a identificar las características de la evaluación docente, desde sus planteamientos en las políticas como en el desarrollo de su implementación en las distintas organizaciones escolares. De tal manera que una revisión periódica de los informes disponibles, sobre todo ante el crecimiento del número de información acumulada, puede constituirse en una plataforma ideal para orientar las acciones futuras de todos los actores involucrados.

En el primer texto se presentan las perspectivas más sobresalientes, resultantes de la revisión de tres décadas de investigación de la evaluación docente, mientras que en los otros dos trabajos se abordan de manera más detallada dos de las perspectivas más prometedoras para progresar en el perfeccionamiento de la evaluación docente: la meta-evaluación de un sistema institucional y la consideración del contexto en la evaluación de competencias profesionales.

Palabras clave: *Investigación de la evaluación docente, meta-evaluación docente, evaluación de competencias, evaluación de desempeño, contexto y evaluación, modelos cognitivos.*

Semblanza de los participantes en el simposio

Nombre del coordinador: Mario Rueda Beltrán

Ex director e investigador Titular del IISUE-UNAM, miembro de la Academia Mexicana de Ciencias e Investigador Nacional del Sistema Nacional de Investigadores Nivel III. Doctor en Ciencias de la Educación en la Universidad de París VIII (Francia) y licenciado en Psicología en la UNAM. Co-coordinador de la Red Iberoamericana de Investigadores sobre Evaluación de la Docencia. Ex director y miembro de comités de revistas especializadas en educación e investigación educativa; ex presidente del COMIE.

Nombre Participante 1: Alejandro Canales

Investigador titular en el Instituto de Investigaciones sobre la Universidad y la Educación, y Miembro del Programa Universitario de Estudios sobre la Educación Superior, ambos de la Universidad Nacional Autónoma de México. Líneas de investigación: evaluación educativa; política educativa; educación superior; y política científica y tecnológica. Doctor en ciencias sociales por la Facultad Latinoamericana de Ciencias Sociales, sede México; Maestro en Ciencias por el Departamento de Investigaciones Educativas del CINVESTAV y licenciado en psicología por la Universidad Nacional Autónoma de México.

Nombre Participante 2: Edna Luna Serrano

Licenciada y maestra en psicología por la Universidad Nacional Autónoma de México y doctora en educación por la Universidad Autónoma de Sinaloa. Investigadora en el Instituto de Investigación y Desarrollo Educativo de la Universidad Autónoma de Baja California (México). Su línea de investigación es evaluación educativa, en particular evaluación de la docencia y evaluación de la formación profesional. Cuenta con una amplia producción científica. Ha coordinado proyectos nacionales y participado como responsable por México en proyectos financiados por la Unión Europea. Es miembro del Sistema Nacional de Investigadores Nivel II, Es directora editorial de la Revista Electrónica de Investigación Educativa.

Nombre Participante 3: Dra. Yolanda Edith Leyva Barajas

Doctora en Educación por la Universidad Autónoma de Aguascalientes, Maestra en Análisis Experimental de la Conducta y Licenciada en Psicología por la UNAM. Actualmente es Coordinadora de Evaluación en el Centro de Estudios sobre la Enseñanza y el Aprendizaje del Derecho, A. C., y recientemente fue Directora General para la Evaluación de Docentes y Directivos en el extinto Instituto Nacional para la Evaluación de la Educación. Es miembro de la Red Iberoamericana de Investigadores sobre Evaluación de la Docencia. Sus principales líneas de investigación se vinculan con la evaluación de competencias profesionales y el efecto del contexto en la práctica profesional de docentes y de profesionales en el ámbito de la salud y del derecho.

Textos del simposio

Balance y perspectivas sobre la evaluación de la docencia

Alejandro Canales y Mario Rueda

La evaluación del desempeño docente sigue siendo un gran reto si se reconoce la diversidad de perfiles del profesorado, los contextos socioculturales y económicos de cada país, las modalidades y los niveles educativos; las condiciones de infraestructura y equipamiento de los espacios escolares, así como las diversas condiciones laborales en las que se ejerce la docencia.

Si bien existen avances en la evaluación docente, como la diversificación de las estrategias, el manejo técnico de los cuestionarios y los aspectos éticos involucrados, aún falta mucho por explorar para que esta actividad contribuya, de forma significativa a su perfeccionamiento. En cuanto la diversificación de estrategias se pueden identificar, iniciativas que han explorado el uso de portafolios, la autoevaluación, la evaluación de pares y del jefe directo, la evaluación formativa en la supervisión clínica, el autoanálisis, los elementos conceptuales y metodológicos para la determinación de estándares, un modelo de evaluación de competencias docentes, así como un modelo y marco guía para la enseñanza en línea. También se constatan intentos por consolidar técnicamente el empleo de los cuestionarios con aportaciones metodológicas para su validez y confiabilidad, así como el estudio de la efectividad docente a partir de las variables asociadas a las características del profesor y del curso. Todas estas tentativas, han sido muy valiosas y han enriquecido el conocimiento del campo de la evaluación y se han constituido en guías de la evaluación de la docencia en múltiples contextos.

La investigación sobre la evaluación de la docencia, orientada a contribuir a la mejora de la enseñanza y el aprendizaje, ha dirigido sus esfuerzos a conocer el estado que guarda el tema en una institución, en una región o en un país; otros trabajos han puesto el interés en la realización de diagnósticos para caracterizar las prácticas de evaluación e identificar iniciativas para mejorarlas. Otros académicos, han desarrollado sus esfuerzos en una visión crítica de los medios habituales para evaluar a los profesores, y en la búsqueda de nuevas formas para valorar su actividad y observar el impacto en la mejora del aprendizaje estudiantil.

También sobresalen, el estudio del contexto institucional para conocer las condiciones que obstaculizan o favorecen el desarrollo de la docencia, la reiteración de utilizar diferentes fuentes de información en los programas de valoración de esta actividad, así como la condición indispensable del involucramiento de los directivos, los profesores y los estudiantes en el diseño y aplicación de estos programas. Un contexto que ha sido relevante para lo que ha ocurrido en las tres últimas décadas y para lo que vendrá.

1. Un breve balance

Imaginar la evaluación del desempeño docente en el futuro, implica comenzar por hacer un breve balance de sus características dominantes actuales en el pasado reciente. La percepción más general y amplia, apunta

hacia reconocer la influencia de la psicología, en particular la corriente promotora del enfoque de la “enseñanza eficaz”, ya que continúa su predominio respecto al modelo de enseñanza en los recintos escolares, reflejado en la presencia actual mayoritaria de los cuestionarios aplicados a los estudiantes.

A finales de los años ochenta, es identificable en el país un conjunto de políticas dirigidas, de acuerdo con el discurso oficial gubernamental, a elevar la calidad de las instituciones de educación superior. Particularmente, destaca la presencia de la evaluación como un recurso para lograr dicha aspiración, en concordancia con las directrices señaladas por algunos organismos internacionales. Las distintas acciones de evaluación ya han sido objeto de análisis particulares para ver en qué medida se lograron los propósitos, e identificar los principales problemas detectados durante su aplicación. Algunos de ellos han mostrado las deficiencias y obstáculos de los programas de evaluación, así como el escaso aprovechamiento de toda la información acumulada. Otra argumentación ha evidenciado la falta de reconocimiento de la diversidad de instituciones que conforman el sistema, en contraste con la inflexibilidad de las iniciativas de evaluación propuestas, agregándose a esto múltiples testimonios de los efectos no deseados registrados en su aplicación (De Vries, y Álvarez, 2015; ANUIES, 2018). Inclusive, los actuales responsables de la dirección del sector educativo han vuelto a plantear la necesidad de reorientar los programas de estímulos a la productividad para que estén acordes con la misión de cada institución, y se otorguen incentivos y reconocimiento a la labor docente (SEP, 2020).

En un seguimiento de los informes presidenciales sexenales de los últimos tres periodos de gobierno, se advierte una ausencia de políticas directamente dirigidas a valorar la función docente en las instituciones de educación superior. Una iniciativa retomada desde finales de los noventa y continuada a lo largo de los sexenios referidos, dio origen a la creación de un Programa del Mejoramiento del Profesorado (PROMEP), para promover la adquisición de posgrados de los académicos de tiempo completo, bajo el supuesto de repercutir en un mejor desempeño e innovación de la docencia. La iniciativa buscaba elevar la calidad de la educación a través de aumentar grados y lograr una combinación de las funciones de investigación y docencia. Una estrategia particular fue lo que se denominó “Cuerpos Académicos”, una agrupación de profesores de tiempo completo (un mínimo de tres) que compartieran una o más líneas de estudio y establecieran metas en común; la agrupación de profesores podían transitar de ser considerados en una etapa inicial de “Formación” a una final de “Consolidación”.

A nivel de investigación académica, en una orientación distinta de las políticas gubernamentales dirigidas a la docencia, la evaluación del desempeño docente ha sido motivo de interés para su tratamiento como objeto de estudio por su creciente presencia en el sistema de educación superior. Las primeras iniciativas de evaluación docente en la universidad, se consignan hacia finales de la década de los setenta y principios de los ochenta, oficialmente hechas para conocer las necesidades institucionales para obtener información sobre la calidad de la enseñanza y las estrategias para la formación de los profesores (Luna y Torquemada, 2008). Sin embargo, es hasta 1996 cuando se identifican las publicaciones de manera continua en evaluación de la docencia presencial, y a partir del año 2000 se hace más evidente la presencia de trabajos sistemáticos sobre este tema.

En un esfuerzo colectivo de documentación, se dio cuenta de los trabajos desarrollados en torno a la evaluación docente en el nivel universitario en la década de los años noventa en el país (cfr. Arbesú, et al. 2006). En el conjunto de textos identificados, poco menos de un centenar, en términos globales se reportó la existencia de materiales en los que predominaban en mayor medida los estudios empíricos, con un soporte de investigación sistemática, seguidos por los trabajos analíticos como resultado principalmente del examen documental y únicamente dos expresando posturas personales. Sobresale la producción sobre temas como los cuestionarios contestados por los estudiantes para valorar la actividad de sus profesores, así como en el tratamiento de los aspectos teóricos de la docencia y su evaluación, las prácticas institucionales sobre ésta y algunas propuestas de nuevos modelos de evaluación. También se identifican materiales que explícitamente evalúan la docencia y la vinculan con algún programa de compensación salarial, por considerar que el impulso de la evaluación, está claramente asociado con los programas gubernamentales.

En el mismo estado de conocimiento se analizan los documentos que dan cuenta de tres tendencias presentes en la vinculación de la práctica educativa y la evaluación de la docencia. La primera está referida a la evaluación de la docencia en el contexto del aula; la segunda a la relación entre ésta y los procesos de pensamiento del profesor y los estudiantes; y la última que intenta el examen completo de la práctica educativa y la evaluación docente. En una última sección se agrupan los trabajos que presentan la vinculación entre las iniciativas de evaluación docente con acciones de formación del profesorado.

Dentro de las características más sobresalientes del conjunto de documentos analizados se destacan las políticas gubernamentales como promotoras de la evaluación de los académicos universitarios. A su vez, no se identificaron iniciativas para procurar la mejora de esta actividad. También se identifica a la evaluación docente como medio para acceder a los programas compensatorios y para cubrir algunos requerimientos de corte administrativo. Por otra parte, los cuestionarios de opinión de los estudiantes, a pesar de múltiples señalamientos de sus limitaciones y problemas técnicos en su diseño y aplicación, aparecen como la forma dominante de atender la solicitud de la evaluación del profesorado. Queda manifiesta la reducida participación de los docentes en el diseño de los programas de evaluación, sobre todo cuando se observa una resistencia o una clara oposición a ellos. La sección dedicada al análisis de los trabajos de la práctica docente, dejaron patente su potencial para desvelar la visión que tienen los propios profesores, y los estudiantes, respecto a la enseñanza, el aprendizaje y la evaluación, en donde los primeros destacaron como los gestores más apropiados para el análisis y la evaluación de su propia práctica y como profesionales reflexivos. Sin embargo, este tipo de ejercicios se identificaron sólo como iniciativas locales que no han sido retomadas para guiar los esfuerzos institucionales.

Una de las recomendaciones que sobresale en la serie de trabajos es la idea de fortalecer la vinculación entre evaluación y formación docente, así como la modificación de aspectos institucionales que puedan reforzar los cambios que a nivel individual se logren como resultado de la participación en las iniciativas de evaluación; así

mismo considerar la posibilidad de trascender la visión sumativa hacia otra más cercana a la retroalimentación y mejora del objeto evaluado.

En una revisión similar de la investigación sobre evaluación de la docencia, ocurrida en el país, pero ahora referida al periodo 2002-2012 (cfr. Luna et al. 2013), se identificaron un total de 168 trabajos. En este esfuerzo por dar cuenta de la producción investigativa en el nivel de educación superior, se agruparon para su análisis en los siguientes rubros: evaluación de la práctica docente en el aula; las políticas de evaluación y compensación salarial; la evaluación y formación, y el análisis de la práctica educativa.

Para el primer rubro, se reportaron la mayor cantidad de trabajos dedicados a estudiar modelos y estrategias de evaluación, en los que se fortalece la búsqueda de diferentes instrumentos y estrategias para su desarrollo. También se formulan propuestas de evaluación que toman en cuenta las distintas características pedagógicas de la licenciatura o el posgrado e igualmente se reportan experiencias de evaluación en diferentes contextos universitarios, o se diseña y valida un modelo completo que sugiere emplear una variedad de recursos. No faltan las aportaciones cuyo interés central es el señalamiento del papel de la teoría en la evaluación o llaman la atención sobre la necesidad de desarrollarla. Adicionalmente otros trabajos analizan los procesos de evaluación docente en diferentes instituciones del país con el propósito de conocer sus características y proponer orientaciones para su perfeccionamiento, además de que, por primera vez, se da a conocer una experiencia de meta evaluación del proceso de evaluación docente en el contexto nacional.

En cuanto los documentos que abordan el tema de las políticas de evaluación y compensación salarial, se reconoce la existencia de una menor cantidad de ellos en relación al conjunto. Entre los principales tópicos tratados destaca la importancia de las condiciones contextuales y las políticas para el éxito de las iniciativas de evaluación, así como la necesidad de clarificar con mayor fuerza la relación entre éstas con los estímulos salariales. Aunque se contempla la relevancia de la docencia y su evaluación, es notoria la ausencia del profesorado en los procesos implicados en ésta y las evidencias de la evaluación con fines de mejora.

La otra categorización empleada, el análisis de la práctica educativa, agrupa al conjunto de trabajos que manejan una conceptualización de la docencia como una actividad compleja y dinámica que posibilita resaltar la exigencia del docente y el grado de compromiso del estudiante, la especificidad de la evaluación en distintos contextos disciplinares, el alejamiento entre el conocimiento pedagógico del docente y su práctica, así como contrastar las percepciones de estudiantes y profesores para identificar un apropiado desempeño en el aula o sus opiniones respecto a los instrumentos de evaluación.

Finalmente, el tratamiento de la sección sobre la relación entre evaluación y formación, advierte sobre el uso de los resultados de los cuestionarios, para considerarlos como parte del diagnóstico de las necesidades de formación. Así mismo, se plantea una propuesta para el desarrollo y evaluación de competencias docentes, en la que se diseñaron dispositivos relacionados con la formación y la evaluación, incluyendo la autoevaluación, la coevaluación, las rúbricas, las exhibiciones y los incidentes críticos.

Los trabajos brevemente reseñados ofrecen un panorama sucinto de la investigación sobre el tema de la evaluación docente en México a lo largo de estas dos décadas, y posibilitan la incorporación de otras personas a esta línea de trabajo.

En otro sentido, en el conjunto de trabajos analizados se pueden distinguir distintos momentos, en el primero de ellos se observa un interés por dar seguimiento al efecto de las políticas gubernamentales. Una vez conocido cómo fue la instauración de las políticas en las instituciones, particularmente las referidas a la evaluación docente, se pudieron caracterizar las prácticas con sus aciertos y limitaciones. En otro momento, la identificación de los recursos y métodos utilizados. De forma reactiva, también se sitúan esfuerzos para abocarse al diseño y experimentación de otras formas de dar cumplimiento a dichas políticas.

También, en un tercer momento, el reconocimiento de escasos efectos positivos en el mejoramiento de la actividad docente, condujo a un grupo de académicos a estudiar las condiciones institucionales que pueden obstaculizar o promover el desarrollo de la docencia. Así, se propuso un modelo (Canales, Leyva, Luna y Rueda, 2014) y se puso en práctica (Rueda, 2016), en el que se consideraban la presencia de influencias externas a las instituciones, las políticas internas a las organizaciones y las condiciones de las aulas en donde tenía lugar la acción docente, como elementos que podrían ofrecer un panorama distinto del quehacer docente. Con este planteamiento se dio un viraje de concebir al maestro como el único agente con la responsabilidad del cumplimiento de la función docente de las instituciones, a una mirada que involucra distintos niveles de influencia e intervención.

2. Expectativas sobre evaluación docente

Si se acepta que a través de la evaluación se escoge y alienta un prototipo de ser docente, podría afirmarse que el tipo de evaluación dominante de la docencia ha fortalecido la idea de que sólo puede haber un modelo apropiado de ejercer la docencia. Más aún, se suscribe de manera implícita que la complejidad de esta actividad puede analizarse a través de los estudios que relacionan los compartimientos específicos del maestro con el aprendizaje efectivo de sus estudiantes. Es decir, el efecto unidireccional de la acción docente sobre el aprendizaje de los estudiantes, por lo tanto el reforzamiento de los roles activo-pasivo en las interacciones en el aula. Todo ello apunta a consolidar, en última instancia, la forma tradicional de concebir y animar el cumplimiento de la enseñanza y el aprendizaje en las instituciones escolares.

De la misma manera, otro elemento a considerar en el devenir de la evaluación docente es la evolución de los estudios sobre la relación educativa que han ido señalando la incorporación de muy diversos componentes en la búsqueda de elementos comprensivos del quehacer en las aulas. Tal es el caso, por ejemplo, de la investigación sobre el impacto de los procesos de pensamiento del profesor y los estudiantes (Shulman, 1989). En este caso, se han desarrollado estudios interesados en desentrañar el pensamiento de los profesores y su influencia en la experiencia escolar, así como el conocimiento de las percepciones de los profesores y estudiantes en la

interacción en la escuela y su impacto en el aprendizaje. En otros trabajos, se ha dado testimonio de la presencia de la cultura dominante en las escuelas ubicadas en comunidades originarias y su desencuentro en las experiencias cotidianas en las aulas (Bertely, 2000). Así mismo, se ha destacado el papel activo de los estudiantes en la construcción del conocimiento, y la presencia de lo ideológico en la vida cotidiana en las escuelas, ejemplificado en el descubrimiento del “currículo oculto”. Todas estas búsquedas coinciden en superar un enfoque simplista que propone entender y explicar los procesos de enseñanza-aprendizaje como la resultante de la correlación de variables causa-efecto, y por el contrario, abonan al reconocimiento de este proceso, concibiéndolo de forma más holística y rescatando su complejidad.

Otro elemento que ha modificado la percepción y las exigencias hacia los sistemas educativos, son los cambios abruptos ocurridos en las sociedades, derivados del agotamiento de los modelos económicos dominantes que han generado una distribución inequitativa de la riqueza y el acceso a los bienes simbólicos para un reducido porcentaje de la población general. Esta situación, ha favorecido el surgimiento de señalamientos críticos hacia las instituciones educativas y a una mayor exigencia respecto a sus posibles funciones a cumplir, en las sociedades contemporáneas. Derivado de lo anterior, también ha surgido el cuestionamiento del papel fundamental que pueden jugar los profesores en los procesos de formación de las nuevas generaciones. Nuevamente, la consideración de este momento actual deberá afectar la forma de acercarse a la evaluación docente del futuro.

Adicionalmente, otro factor a tomar en cuenta al perfilar la evaluación futura, es el reconocimiento de la irrupción actual de las tecnologías de la información y la comunicación en la sociedad en general y, también, a su fuerte presencia deliberada o no planeada, en las instituciones educativas de todos los niveles escolares. El reconocimiento de esta nueva condición debe nutrir la discusión sobre el papel social de la escuela en general, la dinámica de la vida en las aulas, el aprendizaje y los nuevos roles de maestros y estudiantes que deberán tomarse en consideración para la valoración de la labor docente en estos nuevos contextos. Una mirada más holística de lo educativo, como lo propone la teoría ecológica, deberá reconocer el papel de las tecnologías de la información y la comunicación en una sociedad que aspira a ser considerada como la sociedad de la información y el conocimiento. La integración de múltiples aspectos en los procesos de aprendizaje, como los recursos personales y materiales de los estudiantes, los contextos específicos en los que se desarrolla, la integración para el aprendizaje de la motivación y el interés por aprender de los participantes, así como las estrategias empleadas. Todo este conjunto de elementos repercutirá en la manera de abordar la formación inicial y continua de los docentes, y en consecuencia en las formas de evaluar su actividad.

Por todo lo anteriormente planteado, se presenta como necesario continuar la búsqueda por aumentar la comprensión y orientación de esta actividad humana, compleja e importante, que es la docencia. Sobre todo, en un contexto social cambiante, en donde la tecnología ocupa un papel central que moviliza todas las concepciones disponibles de lo que son el aprendizaje y la enseñanza, del papel de la institución escolar, así

como de los correspondientes roles de los profesores y los estudiantes. Por lo que pareciera conveniente cultivar una postura holística en el conocimiento y el diseño de acciones de intervención en el campo de la educación, particularmente en los procesos de enseñanza, el aprendizaje y la formación profesional de los futuros maestros. Desde otro ángulo, por la presencia de las tecnologías de la información y la comunicación, cada vez se hacen más imperceptibles las diferencias entre lo que se conoce como la educación presencial y la clasificada como a distancia, así como entre la educación formal e informal, lo que se sumará también a las nuevas exigencias hacia los sistemas de evaluación docente. Si se amplía más el reto de la evaluación del futuro, habría que pensar en la presencia de la internet en la construcción del conocimiento, en donde éste se genera y difunde en espacios compartidos y colaborativos al grado de pensar en un aprendizaje permanente en donde el aprendiz es quien decide que es lo que quiere aprender, en dónde, cómo y cuando lo hará.

A pesar de varias décadas de investigación sobre la actividad de los docentes, no hay indicadores concluyentes que señalen la mejora de esta actividad como resultado de los programas de evaluación. La importancia y la complejidad de la educación, en particular de la función de la docencia, sustentan la pertinencia de foros de conocimiento, discusión e intercambio de ideas para nutrir y ampliar la comprensión y la búsqueda de las condiciones óptimas para el desarrollo de los Sistemas Educativos. Una de las pocas certidumbres compartidas universalmente es el reconocimiento otorgado a la educación como herramienta clave en la construcción de mejores sociedades. Sin embargo, también siempre están presentes las grandes diferencias de opinión cuando se deben traducir las directrices generales, en propósitos concretos, la selección de medios o recursos para lograrlos, o se diverge en las condiciones o puntos de partida, o en el reconocimiento de la incertidumbre provocada por una sociedad en evolución constante.

Por la diversidad existente de puntos de vista es estratégico favorecer espacios para la reflexión en torno a los fines, los medios y los logros de la educación en nuestros días y en nuestros diversos contextos, y precisar el papel de la docencia y su evaluación. Es necesario, y siempre será bienvenido, aplicar el pensamiento a la educación, y educarse para potenciar las capacidades analíticas de cada uno y de todos. Sólo un pueblo educado puede llevar a cabo juicios sobre su proceder, sus fines y sus logros educativos. Se requiere de más y mejor educación para hacer críticas pertinentes y constructivas.

Los retos que tenemos por delante son de una magnitud y complejidad que requerirán de todos nuestros esfuerzos y capacidades para enfrentarlos, la investigación educativa podrá contribuir a consolidar el conocimiento y orientar nuestras búsquedas hacia mejores escenarios.

Bibliografía

Arbesú, I., Díaz-Barriga, F., Elizalde, L., Luna, E., Rigo, M.A., Rueda, M., y Torquemada, A. D. (2006). *La Evaluación de la Docencia Universitaria en México: Un Estado de Conocimiento del período 1990 -2004*. Perspectiva Educacional, Formación de Profesores. Pontificia Universidad Católica de Valparaíso Viña Del Mar, Chile. Núm. 48, pp. 27-58

- ANUIES (2018) *Visión y acción 2030. Propuesta de la ANUIES para renovar la educación superior en México. Diseño y concertación de políticas públicas para impulsar el cambio institucional.* Asociación Nacional de Universidades e Instituciones de Educación Superior. México. 181 pp.
- Canales, A., Leyva, Y., Luna, E., y Rueda, M. (2014). *Análisis de las condiciones contextuales para el desarrollo de la docencia*, en J. C. Rodríguez-Macías (coord.), *Aportaciones metodológicas a la evaluación a gran escala del aprendizaje y su contexto*, Guadalajara, Editorial Universitaria, pp. 15-28.
- De Vries, W. y Álvarez, G (2015), "Can Reform Policies Be Reformed? Analysis of the evaluation of academics in Mexico", en P. Zgaga, U. Teichler, Hans G. Schuetze y A. Wolter (eds.), *Higher Education Reform: Looking back - looking forward*, Fráncfort, Peter Lang, pp. 257-272.
- Luna, E. y Torquemada, A.D. (2008). *Los cuestionarios de evaluación de la docencia por los alumnos: balance y perspectivas de su agenda*. REDIE. Revista Electrónica de Investigación Educativa, pp 1-15.
- Luna, E., Elizalde, L., Torquemada, A., Castro, A., y Cisneros-Cohernour, E. (2013). *Estado del conocimiento sobre la evaluación de la docencia 2002-2012*. Pp. 380-416. En: Maldonado, A. (coordinadora). (2013). *Educación y Ciencias: políticas y producción de conocimiento 2002-2011*. ANUIES-COMIE, México
- Rueda, M. (coord.). (2016). *Prácticas y condiciones institucionales para el desarrollo de la docencia*. Instituto de Investigaciones Sobre la Universidad y la Educación, UNAM. México.
- Secretaría de Educación Pública, SEP. (2020). Programa Sectorial de Educación 2020-2024. Programa Sectorial derivado del Plan Nacional de Desarrollo 2019-2024. https://www.planeacion.sep.gob.mx/Doc/planeacion/mediano_plazo/pse_2020_2024.pdf Publicaciones, 48(1), 143-159. doi:10.30827/publicaciones.v48i1.7334
- Shulman, L.S. (1989). Paradigmas y programas de investigación en el estudio de la enseñanza: una perspectiva contemporánea. En M.C. Wittrock (ed.): *La investigación de la enseñanza, I. Enfoques, teorías y métodos*. Barcelona: Paidós/MEC, 9-91.

Experiencia de metaevaluación de un sistema de evaluación docente. Retos y perspectivas

Edna Luna Serrano

El propósito de este trabajo es presentar la experiencia de metaevaluación del Sistema de Evaluación Docente (SED) de la Universidad Autónoma de Baja California (UABC) y discutir los retos y perspectivas que enfrentan este tipo de sistemas.

La metaevaluación se define como la evaluación de la evaluación (Scriven, 1991), realizada por medio de la revisión y verificación de la evaluación con el propósito de detectar problemas relacionados con un uso inadecuado, dificultades administrativas o errores técnicos (Stufflebeam, 2011).

Por tanto, la metaevaluación se concibe como un procedimiento para describir y juzgar una actividad de evaluación con base en los supuestos que fundamentan una buena evaluación (Stufflebeam, 2011). Al respecto,

la metaevaluación es una forma de evaluación de orden superior y compromete evaluaciones secundarias o terciarias, según las requeridas para identificar fortalezas y áreas de oportunidad. Stake (1975) menciona que la metaevaluación aporta un control de calidad para la evaluación.

En los programas de evaluación docente, la metaevaluación resulta relevante en función de conocer la calidad de las evaluaciones y profundizar en el proceso evaluativo (Stake, 2017). En este sentido, la metaevaluación ofrece una revisión profunda y sistemática de un programa de evaluación para reconocer buenas prácticas y oportunidades de mejora. Esta tarea se propone a través del estudio y verificación de la evaluación para detectar problemas tales como el uso inadecuado de los resultados de la evaluación, dificultades administrativas o errores técnicos (Stufflebeam, 2011).

La UABC es la institución de educación superior pública más importante de Baja California. Se caracteriza por ser descentralizada en cuanto a su oferta educativa. En el segundo período de 2020 ofertó 135 programas de licenciatura y 46 programas de posgrado. En este mismo período la matrícula de estudiantes licenciatura fue de 63,495. Asimismo, cuenta con 1,631 profesores de tiempo completo (26.5% de la planta docente), 72 docentes de medio tiempo (1.1%), 4,471 profesores de asignatura (72.4%), y 3,508 empleados administrativos (UABC, 2018).

En la UABC la evaluación de la docencia se realiza con base en los cuestionarios de evaluación contestados por los estudiantes. El proceso que involucra la aplicación de los cuestionarios, el procesamiento de la información y devolución de los resultados se realiza a través del Sistema de Evaluación Docente (SED). El SED actual se implementó a partir del primer periodo semestral de 2010 (aunque la aplicación generalizada y sistemática de los cuestionarios de evaluación docente por los alumnos data de 1994, fecha en la cual se incorporó el puntaje de los estudiantes como un rubro en el Programa de Estímulos al Personal Académico [PREDEPA]).

Las particularidades a destacar del SED a continuación se describen con base en dimensiones:

Dimensión Características del Instrumento

El SED incluye tres tipos de cuestionarios diferenciados en función de la modalidad de enseñanza. Estos fueron diseñados para valorar la enseñanza de acuerdo con el modelo de la UABC por competencias y han sido sometidos a rigurosos procesos de validación:

- 1) Cuestionario de Evaluación de la Competencia Docente (CECD) de las asignaturas teóricas presenciales. La primera versión data de 2010, misma que se actualizó en 2017 (Luna, Calderón, Caso y Cordero, 2012; Luna y Piñuelas, 2015).
- 2) El Cuestionario de evaluación de la docencia para las asignaturas prácticas.
- 3) El Cuestionario de evaluación de competencia docente en línea (Luna y Hernández, 2020).

Los reportes que se generan para la entrega de los resultados informan sobre: 1) el promedio general de los puntajes del cuestionario; 2) el promedio por dimensión (planeación y gestión del proceso enseñanza-aprendizaje, y conducción y valoración); 3) el promedio de cada reactivo incluido en el cuestionario; 4) la valoración global que el estudiante otorga al profesor; y un espacio para comentarios libres.

Dimensión Administración del Instrumento

Los cuestionarios se aplican de manera sistemática de acuerdo con el calendario escolar una vez hacia el final de cada semestre a toda la población de estudiantes de licenciatura. Su aplicación y procesamiento de la información se realiza de forma computarizada.

Dimensión Uso de la información

Los reportes proveen información a directivos y profesores. Cabe señalar que, desde 1994 el puntaje de evaluación docente se incorporó como un indicador del PREDEPA, mismo que ha variado a través del tiempo. En la versión más reciente representan el 7.5% de la Tabla de Indicadores Académicos o el 4.5% de la evaluación global (Acuerdo Reglamentario, 2018). Además, la evaluación del profesorado es un elemento a cubrir en los procesos de acreditación.

Algunas consideraciones críticas al SED

Cabe precisar que desde su creación el SED no ha sido sometido a ningún proceso sistemático de evaluación, es decir una evaluación de la evaluación. Al mismo tiempo, se identifican observaciones recurrentes como:

En usos de la evaluación, la crítica generalizada al SED de la UABC se relaciona con el escaso o nulo uso formativo de la evaluación. Al mismo tiempo, no se cuenta con estudios sobre el impacto que tienen los resultados de la evaluación docente en los distintos ámbitos de incidencia como: en la vida académica del profesorado, en el mejoramiento de la actividad docente, en orientar acciones de formación, en la toma de decisiones administrativas (por citar las más obvias).

Además, cabe tener presente que un análisis de los Comités Interinstitucionales para la Evaluación de la Educación Superior (2018) se cuestionó que la evaluación del profesorado se realice solo a través de los cuestionarios de evaluación de la docencia por los alumnos y no se utilicen otras estrategias como evaluación colegiada y autoevaluación, también que, a pesar de las evaluaciones algunos maestros no modifican su enseñanza.

El escenario del SED de la UABC muestra el avance en la consolidación de los instrumentos de evaluación docente con base en la opinión de los estudiantes, el desarrollo del sistema que administra la estrategia de evaluación. Sin embargo, la gran asignatura pendiente se relaciona con el uso de los resultados para la mejora de la actividad docente y de la vida académica, y la implementación de diversas estrategias de evaluación que permitan tener una visión más comprensiva de la función docente.

En este contexto el objetivo general de trabajo fue realizar una metaevaluación al SED con el fin de identificar sus fortalezas y los aspectos a mejorar en: el Cuestionario de Evaluación de la Competencia Docente (CECD), el sistema gestor del SED, la administración del CECD, el uso de los resultados y la infraestructura institucional. Además, se plantearon las siguientes preguntas y objetivos específicos:

Preguntas específicas	Objetivos específicos
¿Cuáles son los elementos y procesos que componen al SED?	Identificar las partes constitutivas y procesos que componen al SED.
¿Cuáles son las principales observaciones y sugerencias de mejora de los actores educativos (directivos, docentes y estudiantes) al CECD?	Caracterizar las fortalezas y debilidades de CECD.
¿Cuáles son las principales observaciones y sugerencias de mejora de los actores educativos (directivos, docentes y estudiantes) al sistema gestor del SED?	Delimitar los aspectos positivos y a mejorar del sistema gestor del SED.
¿Cuáles son las opiniones y recomendaciones de mejora al proceso que administra la aplicación del CECD?	Identificar las diversas opiniones sobre el proceso que administra la aplicación del CECD.
¿Cuáles son las opiniones de los actores educativos (directivos, docentes y estudiantes) en relación con el uso de los puntajes del CECD?	Identificar las diversas opiniones y argumentos sobre los usos de los puntajes del CECD.
¿El SED cuenta con la infraestructura institucional necesaria para su operación?	Valorar la infraestructura institucional disponible para la operación del SED.
¿Cuáles son las innovaciones que proponen los actores educativos?	Caracterizar las innovaciones propuestas al SED

Método

Este estudio corresponde a una metaevaluación realizada desde la perspectiva de la evaluación comprensiva (Stake, 1972, 2006, 2011).

Descripción general de los participantes

Se contó con la participación de dos funcionarios de la administración central (el Coordinador General de Formación Básica y el Coordinación General de Planeación y Desarrollo Institucional de la UABC); así como tres grupos de la comunidad educativa: Grupo 1. Directores y subdirectores de las unidades académicas que imparten programas de licenciatura; Grupo 2. Docentes, y Grupo 3. Estudiantes. Las diversas muestras son probabilísticas por autoselección (ver tabla 1). Los datos fueron recabados en el periodo 2019-2.

Tabla 1. Número de participantes por muestra

	Participantes		
	Directores y subdirectores	Docentes	Estudiantes
Total	41	1,431	10,160

Instrumentos

De manera general, los diversos instrumentos utilizados integraron las dimensiones propuestas por la Red Iberoamericana de Investigadores sobre Evaluación de la Docencia (2008) [RIIED] y las consideraciones referidas por Rueda (2010) para el diseño e implementación de sistemas de evaluación docente. A través de los

distintos instrumentos se indagó respecto a las siguientes dimensiones: 1) la estructura y contenido del CECD; 2) el procedimiento de aplicación del CECD; 3) el uso de los resultados de la evaluación; y 4) la valoración del SED. Cada dimensión se orientó hacia un tema en particular en atención a los objetivos propuestos. De esta manera se indagó sobre aspectos relacionados con las dimensiones metodológica-procedimental y uso de los resultados (RIIED, 2008). Los instrumentos utilizados fueron:

1. Un guión de entrevista semiestructurada dirigida para los coordinadores de la administración central.
2. Un cuestionario dirigido a directores y subdirectores.
3. Un cuestionario para los docentes.
4. Un cuestionario orientado a los estudiantes.

Procedimiento

Las entrevistas se realizaron por medio de videollamada en la aplicación de *Hangouts* de Google y los cuestionarios fueron distribuidos por medio del correo electrónico institucional. En primer lugar, se llevaron a cabo las entrevistas a los coordinadores de la administración central, lo cual permitió identificar las principales observaciones que desde este ámbito se tenían hacia el SED, mismas que fueron retomadas en los distintos cuestionarios. Enseguida se aplicaron los cuestionarios a directores y subdirectores, seguidos de los correspondientes a los docentes y estudiantes.

La información recuperada de inicio se analizó con base en dos categorías: Metodológica-Procedimental y Uso de los Resultados. La categoría metodológica-procedimental indica la estructura y contenido de los instrumentos, el procedimiento de aplicación y la generación de reportes (Rueda, 2010); por tanto, se dividió en tres subcategorías: estructura y contenido del cuestionario; procedimiento de aplicación y comunicación de los resultados. La categoría uso de los resultados se refiere a la delimitación de los usos de la evaluación (RIIED, 2008; Rueda, 2010).

El procedimiento de análisis consistió en la revisión sistemática de los todos los datos recuperados de los cuestionarios aplicados a directores y subdirectores, docentes y estudiantes. Para esta tarea, se conformaron tres archivos, cada uno correspondiente a cada grupo de participantes; el *software* empleado fue *Atlas.Ti*. Esta revisión se sustentó en el análisis cualitativo del contenido con un enfoque de desarrollo inductivo-deductivo de categorías (Mayring, 2000). De esta forma, se identificaron códigos para representar los datos; los códigos fueron organizados en las categorías de análisis conforme a su definición y regla de codificación.

Resultados

A continuación, se presentan los comentarios de mayor frecuencia, coincidencia y relevancia en los tres grupos de participantes (directores y subdirectores, docentes y estudiantes) respecto a la categoría Metodológica-Procedimental organizados en las subcategorías que se derivaron (tabla: 2).

Tabla 2. Comentarios de la categoría Metodológica-Procedimental

Grupos de participantes	Subcategoría Estructura y contenido del cuestionario
Directores y subdirectores	<ol style="list-style-type: none"> 1. Contenido del cuestionario: incluir o modificar elementos sobre los conocimientos, ética y responsabilidad. 2. Estructura: redacción y organización. 3. Número de preguntas: extensión del cuestionario.
Docentes	<ol style="list-style-type: none"> 1. Número de preguntas: el cuestionario es extenso. 2. Desconocimiento del contenido del cuestionario. 3. Sin comprensión: los estudiantes no comprenden los conceptos.
Estudiantes	<ol style="list-style-type: none"> 1. Número de preguntas: el cuestionario es extenso. Integrar elementos referentes a la ética. 2. Preguntas inadecuadas: respecto a la asignatura o área de conocimiento. 3. Calidad de las preguntas.
Grupos de participantes	Subcategoría Procedimiento de aplicación
Directores y subdirectores	<ol style="list-style-type: none"> 1. Período de aplicación: horarios específicos por UA; mayor periodo. 2. Condicionamiento: obligatoriedad de la evaluación para reinscripción. 3. Software: indicaciones sobre el funcionamiento deficiente del sistema gestor del CECD (lentitud, sin función o carga).
Docentes	<ol style="list-style-type: none"> 1. Evaluación irreflexiva: los estudiantes generalmente realizan la evaluación sin atender cabalmente lo que se les solicita. 2. Evaluación obligatoria: los estudiantes realizan la evaluación como un requisito para poder reinscribirse o ingresar a la subasta de materias. 3. Evaluación subjetiva (poco confiable). 4. Desinformación de los estudiantes sobre el propósito de la evaluación.
Estudiantes	<ol style="list-style-type: none"> 1. Período de aplicación insuficiente. 2. Software, diseño instruccional: indicaciones sobre el funcionamiento deficiente (lentitud, sin función o carga) y diseño obsoleto (falta de actualización del ambiente virtual) del sistema gestor del CECD. Dificultad de acceso. 3. Evaluación irreflexiva: los estudiantes generalmente realizan la evaluación sin atender cabalmente lo que se les solicita.
Grupos de participantes	Subcategoría Comunicación de los resultados
Directores y subdirectores	<ol style="list-style-type: none"> 1. Medios de entrega para la devolución de los resultados de la evaluación docente (generalmente se realiza cuando se obtiene un bajo nivel). 2. Resultados de evaluación fuera de tiempo. 3. Puntuaciones poco comprensibles o incompletas.
Docentes	<ol style="list-style-type: none"> 1. Desconocimiento total o parcial de los resultados de la evaluación. 2. Resultados de evaluación fuera de tiempo. 3. Desconocimiento para el acceso a los resultados.
Estudiantes	Comunicación de los resultados (solicitud).

Categoría Uso de los Resultados

Esta categoría presenta los principales comentarios sobre el uso de los resultados de la evaluación que manifestaron los directores y subdirectores, docentes y estudiantes de la universidad (ver tabla 3).

Tabla 3. Comentarios principales de la categoría Uso de los Resultados

Grupos de participantes	Categoría Uso de los Resultados
Directores y subdirectores	<ol style="list-style-type: none"> 1. Formación continua: propuestas de formación para los docentes. 2. Implicaciones laborales. 3. Retroalimentar la práctica docente.
Docentes	<ol style="list-style-type: none"> 1. Uso de la evaluación como venganza: los estudiantes usan la evaluación como represalia. 2. Uso de la evaluación para expresarse: los estudiantes se expresan, opinan o se quejan. 3. Uso de la evaluación para mejorar la práctica.
Estudiantes	<ol style="list-style-type: none"> 1. Los resultados no se toman en cuenta. 2. Valoración de la opinión de los estudiantes, se demanda que su opinión sobre el desempeño del docente sea considerada por los tomadores de decisiones. 3. Sin cambio en la docencia.

Síntesis de los resultados

Subcategoría Estructura y contenido del cuestionario

1. Es necesario separar al CECD de Teoría del CEPD-Laboratorio. Son dos cuestionarios independientes dirigidos a asignaturas diametralmente opuestas. Además, la aplicación del CEPD-Laboratorio se hizo de forma generalizada en las asignaturas que tienen algún componente práctico, sin distinguir que este cuestionario solo aplica para la docencia en laboratorio de las áreas de ciencias naturales y salud.
2. Se requiere valorar la pertinencia de incluir reactivos que valoren el componente ético de la docencia.
3. Se precisa realizar la difusión suficiente y oportuna sobre el contenido del CECD.

Subcategoría Procedimiento de aplicación

1. Revisar la pertinencia del periodo de aplicación del CECD.
2. Valorar la eficiencia del sistema gestor del CECD.
3. La falta de credibilidad en el SED alude a la validez interna de la evaluación. Respecto a las observaciones identificadas, se sugiere llevar a cabo una campaña de sensibilización sobre los propósitos de la evaluación y la importancia de una participación responsable por parte de los estudiantes.

Subcategoría Comunicación de los resultados

1. En general, se desconoce en cuánto tiempo el sistema gestor del CECD pone a la disponibilidad de los usuarios los resultados de los puntajes.
2. Los principales usuarios (docentes y directivos) que acceden a los resultados de los puntajes consideraron que se realiza fuera de tiempo, de forma no oportuna su comunicación.

3. Una proporción significativa de la muestra de docentes refirió desconocer las características del sistema gestor del CECD.

Categoría Uso de los Resultados

1. El uso de los resultados por parte de los directivos es poco claro. En los hechos destaca la ausencia de una estrategia institucional orientada a utilizar los resultados para la mejora de la actividad.
2. Los docentes expresaron dos situaciones paradójicas, por un lado, que las respuestas de los estudiantes no son confiables y por otro, que los comentarios les son útiles. Es necesario considerar una estrategia para mejorar la credibilidad de la evaluación.
3. Se requiere reforzar el uso de los resultados con fines de mejora de la actividad docente para mejorar la validez consecuencial.

Discusión de resultados y conclusiones

Este trabajo se propuso realizar una metaevaluación del SED. Los resultados dan cuenta de la complejidad del objeto de estudio y de los retos a asumir a favor del mejoramiento de la calidad de la enseñanza.

En consideración a que la validez interna es la base de la credibilidad de la evaluación, es importante revisar los aspectos que la fortalecen o debilitan. En este caso, se identificaron elementos que la robustecen, como la adecuación entre los referentes teóricos de los cuestionarios de evaluación por los estudiantes con el modelo pedagógico de la universidad; asimismo, los resultados revelaron problemas que atentan contra ella, por ejemplo, en la estructura y contenido del cuestionario es desafortunado el presentar de forma conjunta el CECD de Teoría con el CEPD-Laboratorio (esta situación ha provocado aumentar de manera significativa el número de reactivos que los estudiantes deben contestar, y falta de coherencia entre el contenido del CECD y las asignaturas), además, se percibe la necesidad de incluir otros instrumentos de evaluación formativa.

En relación con la validez externa relacionada con el principio de utilidad de los resultados de la evaluación, las observaciones de los docentes dan cuenta de la subutilización del SED, por un lado, se refirió sobre el desconocimiento del contenido del CECD y los resultados de los puntajes, y por otro, la falta de credibilidad en las respuestas de los estudiantes. Este hecho se refuerza en las contradicciones planteadas por los directivos respecto a la comunicación de los resultados y su uso. Al respecto, se considera necesario implementar una estrategia institucional integral que oriente el uso formativo de los puntajes.

Lo anterior son solo algunos ejemplos de la necesidad de realizar procedimientos de metaevaluación con el fin de regular los procesos para el cumplimiento de los fines para los cuales fueron diseñados, así como, las mejoras acordes a los cambios contextuales propios de la vida académica.

Referencias

- Luna, S. E. Calderón, G. N. Caso, N. J., y Cordero, A. G. (2012). Desarrollo y validación de un cuestionario de evaluación de la competencia docente con base en la opinión de los estudiantes (pp. 119-158). En E. Chacón, B. García-Cabrero, E. Luna y R. Marín (Coords.). *Evaluación de competencias docentes en la educación superior*. México: Juan Pablos Editores.
- Luna, S. E. y Reyes, P. E. (2015). Validación de constructo de un cuestionario de evaluación de la competencia docente. *Revista Electrónica de Investigación Educativa*, 17(3), 13-27. <http://redie.uabc.mx/redie/article/view/1090>
- Luna Serrano, E., y Hernández Villafaña, A. D. (2020). Desarrollo de un cuestionario de evaluación de la competencia docente en línea. RIED. *Revista Iberoamericana de Educación a Distancia*, 23(2), pp. 307-328. <https://dx.doi.org/10.5944/ried.23.2.27084>
- Mayring, P. (2000). Qualitative content analysis. *Forum qualitative social research*, 1(2). <http://www.qualitative-research.net/index.php/fqs/article/view/1089/2385>
- Red Iberoamericana de Investigadores de la Evaluación de la Docencia. (2008). Reflexiones sobre el diseño y puesta en marcha de programas de evaluación de la docencia. *Revista Iberoamericana de Evaluación Educativa*, 1(3), 163-168. <https://revistas.uam.es/index.php/riee/article/view/4661/5098>
- Rueda, M. (2010). Reflexiones generales a considerar en el diseño y puesta en operación de programas de evaluación de la docencia. *Revista Iberoamericana de Evaluación Educativa*, 3(1), 344-350. <https://revistas.uam.es/index.php/riee/article/view/4552/4986>
- Scriven, M. (1991). *Evaluation thesaurus* (4ta. ed.). SAGE.
- Stake, R. E. (1975). *Program evaluation particularly responsive evaluation*. Center for Instructional Research and Curriculum Evaluation, 5. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.200.333&rep=rep1&type=pdf>
- Stake, R. E. (2017). Meta-evaluation of systems for the evaluation of teaching. En E. Luna y M. Rueda, *Experiencias de evaluación de la docencia en Iberoamérica* (pp. 248-256). Universidad Autónoma de Baja California.
- Stufflebeam, D. L. (2001). The Metaevaluation Imperative. *American Journal of Evaluation*, 22(2), 183-209. [https://doi.org/10.1016/S1098-2140\(01\)00127-8](https://doi.org/10.1016/S1098-2140(01)00127-8)
- Stufflebeam, D. L. (2011). Meta-evaluation. *Journal of MultiDisciplinary Evaluation*, 7(15), 99-158. http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/300
- Universidad Autónoma de Baja California. (2017). *Formación docente: Programa Flexible De Formación y Desarrollo Docente (PFFDD)*. <http://www.uabc.mx/formacionbasica/fdocente.html>

El contexto en la evaluación de competencias profesionales: avances conceptuales y metodológicos

Dra. Yolanda E. Leyva Barajas

Introducción

Resultan cada vez más evidentes las brechas entre el conocimiento, habilidades y competencias que adquieren los estudiantes en las escuelas de educación superior y las que demanda la vida real y profesional para la solución de problemas de diversa índole. Si bien a nivel internacional, existen algunos programas cuyo diseño está basado en nuevas pedagogías para facilitar el desarrollo de los estudiantes en la adquisición de competencias como ciudadanía, colaboración, comunicación, creatividad y pensamiento crítico (McEachen, 2017), lo cierto es que, en su mayoría se quedan a un nivel de declaraciones teóricas y de buenas intenciones, por la dificultad que implica implementar modelos de intervención y de evaluación que den cuenta de estos constructos complejos.

La brecha respecto a las necesidades del mundo laboral, usualmente se atiende mediante programas de formación para el trabajo en las instancias públicas y privadas responsables de prestar algún servicio profesional, con resultados que no siempre llegan a cubrir las necesidades de formación de todas las personas, en gran medida por las dificultades técnicas implicadas en el diseño e implementación de modelos de evaluación e intervención apropiados, lo que dificulta contar con programas de desarrollo profesional que logren satisfacer las necesidades detectadas.

En esta ponencia nos centramos en la revisión de algunos avances de cómo evaluar competencias profesionales dentro de una perspectiva actualizada con una base conceptual que permita la creación de nuevas medidas de progreso y ejecución. Mientras que las medidas tradicionales reducen al aprendizaje dentro de los confines de los que las personas evaluadas saben, el foco cambia su enfoque del conocimiento en sí, hacia requerimientos centrados en la capacidad para construir soluciones y tomar decisiones a partir de análisis de los diversos contextos en los que ocurre la práctica profesional.

Es importante mencionar que, en las últimas tres décadas, la evaluación de competencias y capacidades, se ha identificado como una prioridad no solo de los sistemas educativos, sino del mundo laboral. Esto ha impulsado por una parte el uso de modelos cognitivos en el desarrollo de los marcos teóricos de las evaluaciones, y el desarrollo de métodos para asegurar la calidad, diversidad y exhaustividad de la evidencia de evaluación que se recoge. Ambos aspectos han coadyuvado a la medición de lo que desde la teoría en ciencias cognitivas se ha denominado como el aprendizaje profundo (McEachen (2017).

Desde este enfoque teórico, se aborda el tema de la evaluación a partir de extender la discusión hacia un enfoque explícito del rol de la evaluación, es decir cómo se usa y cómo sirve al aprendizaje profundo y al ejercicio profesional. La medición y evaluación educativa ha adquirido un nuevo y dinámico significado desde esta

perspectiva, aunque también ha representado desafíos importantes en los dos ejes de trabajo de la evaluación: 1) la construcción de marcos teóricos para interpretación de los resultados y 2) el diseño y desarrollo de diversos tipos de tareas que pongan en juego procesos mentales de alto orden implicados en la transferencia de conocimiento a una diversidad de contextos, para recabar información útil para la emisión de juicios y la toma de decisiones en términos de mejora del ejercicio profesional.

Desarrollos recientes en la evaluación de competencias profesionales

Entre los desarrollos más recientes e importantes en la evaluación, tenemos la construcción de marcos teóricos basados en modelo cognitivos, como soporte para llevar a cabo inferencias acerca de procesos y estructuras mentales generalmente asociados a la toma de decisiones para la solución de casos y problemas que se presentan en lo general en cualquier campo profesional. Un modelo cognitivo consiste en una explicación teórica de los procesos y pasos que asumimos ocurren durante un fenómeno cognitivo complejo, como la solución de casos o problemas, la toma de decisiones, la planeación o incluso la recuperación de información de la memoria humana (Busemeyer & Diederich, 2010; Markman, 1999).

El desarrollo de un modelo cognitivo es un intento formal para caracterizar estos procesos, lo cual permite hacer supuestos particulares acerca de los mecanismos involucrados en la ejecución de tareas específicas dentro de un conjunto particular de condiciones. En el campo de la evaluación de constructos complejos, la lógica que subyace es que un modelo cognitivo puede proporcionar a priori, supuestos y principios válidos para el diseño, medición y validación de diversas tareas evaluativas requeridas en una evaluación de competencias (Busemeyer & Diederich, 2010).

Un marco teórico robusto es la base para la definición del dominio a evaluar, el cual consiste en hacer un muestreo representativo de las observaciones que se requieren -expectativas de desempeño esperado-, que apoyen las inferencias en torno a los procesos mentales asociados a las competencias profesionales a evaluar. El foco de un constructo complejo, implica la evaluación de competencias asociadas a la habilidad de aplicar el conocimiento de contenidos teóricos o prácticos en situaciones que entrañan ciertos márgenes de complejidad técnica.

Esta combinación hace que el constructo sea multidimensional en un sentido que implica que no refleja el contenido de conocimiento aislado, sino la habilidad requerida para aplicarlo en una diversidad de situaciones y contextos, lo cual representa una diferencia fundamental entre un constructo simple y uno complejo con implicaciones importantes para el diseño y desarrollo de evaluaciones válidas. Metodologías basadas en solución de casos, se han implementado como una alternativa prometedora para la evaluación de competencias porque permiten valorar la capacidad de considerar cómo el contexto afecta el caso o problema a enfrentar (Leyva, 2011). Con ello se está dando cuenta de la capacidad de los profesionales para transferir conocimiento útil a diversos contextos.

La evidencia acumulada de la investigación educativa en el campo de la evaluación de competencias mediante el uso de modelos cognitivos, ha sido útil principalmente para los procesos de construcción de rúbricas de calificación con el propósito de estandarizar los juicios de las personas encargadas de valorar el desempeño esperado. Los dos tipos de rúbricas más utilizadas en la evaluación de desempeño son las analíticas y las holísticas, las cuales difieren en términos de su impacto, poder de discriminación, confiabilidad inter-jueces, grado de sesgo y en cuanto a eficiencia costo-beneficio -tiempo, esfuerzo y dinero- (Kuo, 2007). La elección de una u otra es significativa porque representa de manera explícita o implícita las bases teóricas de la evaluación (Weigle, 2002).

La importancia de estos desarrollos, además de que permiten mejorar los procesos de calificación, constituyen un insumo fundamental para llevar a cabo procesos de retroalimentación del desempeño observado, basados en los procesos implicados en la toma de decisiones y la solución de casos o problemas, lo que permite proporcionar información más detallada de los aspectos a mejorar. En esto reside la utilidad de la evaluación porque permite apoyar la toma de decisiones a nivel individual de las personas evaluadas. También ha derivado en la detección de áreas de oportunidad más generales de grupos o subgrupos de personas evaluadas, con lo cual, se facilita la toma de decisiones a nivel institucional, para orientar mejor los recursos de formación continua del personal que constituye el capital humano de sus organizaciones laborales.

Evaluación de competencias a partir de solución de casos

Debido a la larga historia de la evaluación basada en competencias en la medicina, se ha acumulado considerable evidencia acerca de las características psicométricas de los métodos de evaluación empleados. Leyva (2011) sintetiza esta evidencia en una serie de lecciones aprendidas en el contexto de la evaluación de las ciencias de la salud. Esta información, si bien se ha obtenido en un campo profesional determinado, aporta información útil para el desarrollo y validación de tareas evaluativas centradas en casos reales. Por ello se han venido utilizado en otros campos profesionales, principalmente por lo que esta información aporta para el diseño y desarrollo de tareas evaluativas en pruebas prácticas de desempeño profesional.

Sin embargo, el hecho de que las personas sean evaluadas en situaciones reales de ejecución, no garantiza por sí mismo que el diseño de la prueba y el esquema de muestreo sean apropiados. El muestreo debe considerar tanto el contexto (situación / tarea) como el constructo (conocimiento / habilidad) y las interacciones complejas que se presentan entre estas dimensiones. Todo esto forma parte de la definición del dominio a evaluar, lo que normalmente se representa en un matriz de muestreo.

Aunado a la relevancia de una buena definición del dominio a evaluar, la evaluación en la actualidad tiene que cumplir con un propósito formativo, independientemente de que esté ligada a otros propósitos más formales de acreditación o certificación, por lo que resulta fundamental entender la importancia de la dimensionalidad de un constructo complejo -en el contexto psicométrico- para diseñar evaluaciones válidas, confiables y útiles (Gorin, 2006).

A nivel técnico, una vez definido el dominio a evaluar, las expectativas de desempeño se pueden capturar bajo los principios de la evaluación auténtica que usa métodos mixtos, la cual incluye un amplio rango de indicadores para revelar no solo lo que las personas evaluadas conocen, sino lo que hacen, lo que pueden hacer y cómo pueden continuar mejorando su desempeño. Se requiere de un amplio rango de tareas evaluativas que se incrusten en experiencias de aprendizaje profundo para lograr mediciones robustas significativas. Recientemente, el portafolios de evaluación ha sido la respuesta a la integración de una gran variedad de evidencia asociada a las competencias para el ejercicio profesional. El uso del portafolios desde los noventas, se ha incorporado en el ámbito educativo y laboral como una estrategia que ha tenido buenos resultados en términos de utilidad de la evaluación y para promover procesos de reflexión (Barbera, 2005; Leyva, 2018) que favorecen el aprendizaje profundo y con ello la posibilidad de transferir conocimiento útil a una diversidad de contextos.

No obstante, la diversidad de evidencia que se requiere en la evaluación de competencias, ha representado nuevos desafíos para los procesos de calificación; como la de integración de una puntuación que refleje adecuadamente el desempeño esperado en las respuestas de las personas evaluadas. Al respecto, también existen desarrollos importantes que ayudan a la definición de un modelo de calificación que permita emitir un dictamen sobre las competencias evaluadas con base en los usos previstos de la evaluación. En este ámbito existen desarrollos en dos vertientes: 1) Desarrollos recientes en métodos para el establecimiento de puntos de corte (Cizek y Bunch, 2007) y 2) Desarrollo de diversos modelos de calificación, desde aquellos son conjuntivos, compensatorios, disyuntivos y posibles combinaciones que se pueden definir con base en los objetivos de la evaluación (Martínez, Schweig, & Goldschmidt, 2016).

La investigación en ciencias cognoscitivas también proporciona algunas herramientas analíticas de mucha utilidad, desde una perspectiva muy diferente a aquellas derivadas de la aproximación psicométrica tradicional. En el nuevo campo de la psicometría, la evaluación de un amplio rango de pruebas que se utilizan actualmente en educación, ha requerido esbozar una teoría capaz de modelar esta diversidad. Algunos autores consideran que la teoría de la generalizabilidad tiene ventajas sobre la teoría clásica en cuanto a la posibilidad de modelar la complejidad de las diversas alternativas en evaluación.

Asimismo, autores como Shavelson y Ruiz-Primo (2000) consideran que se requiere expandir la teoría de validez hacia una teoría que dé cuenta de interpretaciones de procesos cognoscitivos a partir de medidas de una prueba o de ejecuciones observables. Desde esta perspectiva el esquema de muestreo es útil para identificar facetas que caracterizan la medición. Las facetas incluyen: a) la tarea presentada, b) la ocasión de la medición, c) los juicios de quienes observan la ejecución, y d) los métodos de evaluación (Ruiz-Primo y Shavelson, 2001).

Los desarrollos más recientes a partir de esta aproximación se centran en la importancia de la dimensionalidad de un constructo complejo para diseñar evaluaciones. Específicamente, necesitamos entender si los distintos componentes de un constructo complejo, se manifiestan psicométricamente como dimensiones latentes separadas, lo que ayudará a comprender cómo se relacionan los componentes de constructos complejos con la

variación en los datos de respuesta de una persona evaluada, conduciendo a interpretaciones más significativas de su ejecución en ambos tipos de evaluaciones.

En términos prácticos, los resultados de estas líneas de investigación, resultan especialmente importantes para que los procesos de desarrollo y validación de evaluaciones de constructos complejos, se centren en la validación de la estructura latente, a partir de un marco conceptual que permita construir argumentos teóricos que después se puedan verificar empíricamente. El diseño y desarrollo operacional de instrumentos de evaluación se debe elaborar a partir de estos argumentos teóricos, y la obtención de evidencia empírica de la validez de estos argumentos. Esta evidencia de validez incluye validez basada en procesos de respuesta y validez interna basada en la estructura latente.

La experiencia reciente de evaluación de competencias profesionales mediante el Portafolios CEEAD

En este trabajo presentamos una experiencia de evaluación criterial de competencias profesionales, desarrollado a partir de un marco teórico basado en un modelo cognitivo general de procesos implicados en la solución de casos reales en diversos contextos. La metodología para el diseño y desarrollo se llevó conforme a los estándares de evaluación publicados por la AERA, APA y NCME (2014). Se llevó a cabo el diseño de un Portafolios de evaluación para obtener evidencia de las competencias implicadas en la solución de casos que atienden lo profesionales del Derecho que se desempeñan como fiscales dentro del Sistema de Justicia Penal Acusatorio (SJPA).

Derivado del marco teórico, se construyó un perfil referencial basado en estándares de competencia profesional, con 4 dimensiones de las cuales dos son transversales a todo el proceso: La atención a los intervinientes y la toma de decisiones estratégicas; y las otras dos corresponden a los dos momentos del proceso penal: La dirección de la investigación y la sustentación de pretensiones ante el órgano jurisdiccional.

El diseño del portafolios, se llevó a cabo a partir del diseño de dos tipos de tareas evaluativas basadas en las dimensiones que ubican etapas específicas: 1) Tareas que demandan la explicación y justificación escrita de acciones y decisiones que se toman para dirigir una investigación y que evalúan aspectos asociados a la planeación de líneas de investigación y la elaboración de una teoría del caso; y 2) tareas de ejecución para evaluar las competencias asociadas a la sustentación de pretensiones ante el órgano jurisdiccional, mediante la observación del desempeño en audiencias preliminares y juicio oral. Mientras que las competencias de las dimensiones transversales a todos los momentos, se evaluaron incluyendo criterios en las rúbricas de calificación en las tareas evaluativas desarrolladas.

En cuanto al formato de respuesta, los sustentantes tienen que: 1) seleccionar un caso que hayan atendido bajo el enfoque actual del SJPA; 2) Hacer la descripción de los hechos del caso, 3) En las tareas escritas, justificar algunas de las decisiones específicas que tomaron en los distintos momentos del proceso, de acuerdo con lo solicitado en cada tarea; o subir un video de una audiencia o juicio oral, de acuerdo a lo solicitado en cada

tarea; y 4) Subir artefactos de su práctica retomados de las carpetas de investigación, como evidencia directa de su desempeño.

El portafolios se aplicó a un primer grupo de 128 fiscales de 5 entidades que operan el SJPA, mediante una plataforma denominada Portafolios CEEAD que consta de dos módulos: el Módulo de aplicación dirigido a las personas evaluadas para que de acuerdo con las tareas evaluativas reconstruyan los casos reales seleccionados y suban los artefactos de su práctica, y el módulo de calificación con las rúbricas analíticas desarrolladas a partir de la definición del dominio a evaluar. La administración fue totalmente en línea y las personas evaluadas pudieron responder desde su hogar o desde las fiscalías en las que trabajan, las cuales proporcionaron espacios con la infraestructura y el equipamiento requeridos para la evaluación. La duración para integrar el portafolios fue de 8 semanas; no obstante, se agregaron dos semanas más de prórroga, en virtud de que la evaluación sucedió en tiempos de pandemia (septiembre a noviembre de 2020).

Para determinar el modelo de calificación, se implementaron dos acciones importantes para definir el modelo de calificación: Una relacionada con el establecimiento de puntos de corte entre los niveles de desempeño esperado, mediante una metodología basada en juicios de especialistas de la familia de los métodos holísticos. Esto constituye una innovación, al menos en el contexto mexicano, ya que en general los métodos holísticos se utilizan principalmente en evaluaciones con fines de certificación de competencias profesionales, lo cual añade valor tanto a los participantes como a quienes otorgan la certificación porque se comunica lo que resulta importante para tener éxito en el área especial que se está examinando (Cizek & Bunch, 2007). De los métodos holísticos existentes, se consideró que el más adecuado a las necesidades y el tipo de evaluación de Portafolios implementado para la evaluación de habilidades de los operarios del SJPA, es el denominado *the body of work* (BoW) desarrollado por Kingston, Kahl, Sweeney & Bay (2001), cuya principal característica consiste en la evaluación de conjuntos completos del trabajo de sustentantes por uno o más jueces que rinden un solo veredicto holístico o juicio global para cada muestra de trabajo.

También se llevó a cabo un análisis de la literatura para construir el modelo de calificación más apropiado a la evaluación y a los usos previstos de los resultados de la evaluación. Entre las decisiones importantes que se tomaron a nivel del diseño y desarrollo de las rúbricas de calificación, fue el hecho de que fueran rúbricas analíticas de 4 niveles de desempeño, en donde el nivel 3 representaba el nivel competente, el nivel 4 un nivel destacado o más consolidado como el de un experto, el nivel 2 representaba un nivel por debajo del desempeño esperado con omisiones o errores que podría afectar el ejercicio profesional, pero cercano al estándar esperado y el nivel 1 más alejado del desempeño esperado. Para esta construcción se utilizó información de estudios entre expertos y novatos como el de Dreyfus y Dreyfus (1980).

Mientras que, para la integración del dictamen global de desempeño, se tomó la decisión de construir un modelo de calificación que considera tanto la puntuación global del portafolios, como la puntuación obtenida en cada tarea evaluativa, principalmente en aquellas que los expertos en el dominio a evaluar, consideraron

esenciales para el ejercicio profesional. En este ejercicio, se tuvo como insumo la información del mapeo de tareas en las matrices de muestreo que representan el dominio a evaluar. Estos procesos permitieron establecer las condiciones asociadas a cada uno de los tres niveles del dictamen global: Destacado, Satisfactorio y Aún no satisfactorio.

Con relación a los resultados de la evaluación, las personas evaluadas recibieron en sus reportes individuales de resultados, junto con el dictamen global y la puntuación obtenida en cada tarea evaluativa, una retroalimentación personalizada a nivel de cada criterio de las rúbricas. La retroalimentación incluyó fortalezas y áreas de oportunidad y se programó de manera automatizada con base en los resultados obtenidos por cada participante. Dado que la puntuación 3 se consideró un desempeño competente y 4 sobresaliente, si la puntuación en el criterio era de 3 o 4 se extraía la información para indicar las fortalezas y si la puntuación era 2 o 1, se extraía la información para la sección de áreas de oportunidad, enfatizando omisiones o errores conceptuales o técnicos a mejorar. Desde luego esto solo es posible si se mantiene la coherencia interna de los instrumentos con base en el marco de interpretación que se construye como parte de la definición del dominio a evaluar con base en el marco teórico que guía la evaluación. También se integró un reporte global y reportes institucionales para las 5 fiscalías participantes con retroalimentación para cada nivel de decisión.

Finalmente, los análisis psicométricos realizados bajo la perspectiva de la Teoría de la Generalizabilidad, nos permitieron no solo corroborar la confiabilidad de los resultados obtenidos, sino que nos permitieron integrar un plan de mejora tanto de los instrumentos como de los procesos de evaluación y calificación para las siguientes ediciones de la evaluación, con estudios de decisión que inclusive nos permiten conocer el efecto que tiene cada tarea en la confiabilidad total del portafolios. Esta información nos ayuda a tomar decisiones en el sentido de disminuir o aumentar el número de tareas evaluativas, y también respecto al efecto del número de evaluadores, los procesos de calificación, el entrenamiento en el uso de las rúbricas, entre otros aspectos relevantes.

Si bien, dentro de esta perspectiva es relevante, tener una agenda de investigación que nos permita obtener evidencia de validez de la evaluación realizada, se puede considerar que los resultados obtenidos son promisorios, y que reflejan las ventajas de haber implementado los avances hasta ahora desarrollados en la evaluación de constructos complejos.

Referencias

- AERA, APA & NCME (2014), *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Barberá, E. (2005): La evaluación de competencias complejas: la práctica del portafolio. *Educere*, vol. 9, núm. 31, pp. 497-504 Universidad de los Andes Mérida, Venezuela.
- Busemeyer, J. R., & Diederich, A. (2010). *Cognitive modeling*. Los Angeles, CA: Sage.
- Cizek, G.J. y M.B. Bunch (2007), *Standard Setting: A guide to establishing and evaluating performance standards on tests*, California, SAGE Publications.

- Dreyfus S. y H. Dreyfus (1980). *A five-stage model of the mental activities involved in directed skill acquisition*. Operations Research Center, University of California, Berkeley.
- Gorin, J. S. (2006a). Item design with cognition in mind. *Educational Measurement: Issues and Practice*, 25(4), 21–35.
- Kingston, N.M. Kahl, S. R., Sweeney K. & Bay L. (2001) Setting performance standards using the Body of Work method. In Cizek, G. J. (Ed.), *Standard setting: Concepts, methods, and perspectives* (pp. 219-248). Mahwah, NJ: Erlbaum
- Kuo, S. (2007). Which rubric is more suitable for NSS liberal studies? Analytic or holistic? *Educational Research Journal*, 22(2), 179-199.
- Leyva, Y. (2011), *Validez de constructo en la evaluación de competencias médicas mediante pruebas referidas a criterio*, Editorial Académica Española.
- Leyva, Y. (2018): Una propuesta para el desarrollo profesional de docentes en México. Red. *Revista de evaluación para docentes y directivos*. Año 4. INEE. <https://www.inee.edu.mx/wp-content/uploads/2019/01/Red11.pdf>
- Markman, A. B. (1999). *Knowledge representation*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Martínez, J. F.; Schweig, J. & Goldschmidt, P. (2016): Approaches for Combining Multiple Measures of Teacher Performance: Reliability, Validity, and Implications for Evaluation Policy. *Educational Evaluation and Policy Analysis*, Vol. 38, No. 4, pp. 738–756
- McEachen, J. (2017). Assessment for Deep Learning. Ontario, Canada: Fullan, M., McEachen, J., & Quinn, J. *New Pedagogies for Deep Learning: A Global Partnership* Retrieved: <http://teachingliterature.pbworks.com/w/file/fetch/135692085/Assessment-for-Deep-Learning.pdf>
- Ruiz-Primo, M.A. y R.J. Shavelson (2001), “Comparison of the Reliability and Validity of Scores from two Concept-Mapping Techniques”, *Journal of Research in Science Teaching*, vol. 38, núm.2, pp. 260-278.
- Shavelson, R.J. y M.A. Ruiz-Primo (2000), On the Psychometrics of Assessing Science Understanding, en J. Mintzes, J. Wandersee y J. Novak (eds.), *Assessing Science Understanding*, (pp. 303-341) San Diego, Academic Press.
- Weigle, S. C. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.