

La validación de un modelo de evaluación de la competencia docente didáctica en una escuela normal.

Validating an evaluation model of didactic teaching competencies in a normal school.

César Augusto Cardeña Ojeda¹

Recibido: 14/10/2022

Aceptado: 13/01/2023

Revista RELEP, Educación y Pedagogía en Latinoamérica.

Disponible en:

<https://iquatroeditores.org/revista/index.php/relep/index>

<https://doi.org/10.46990/relep.2023.5.2.1021>



Resumen

Se planteó el objetivo de validar un conjunto de acciones para evaluar la competencia docente en una escuela normal pública; se siguió un método cuantitativo con la aplicación de fórmulas de inferencia estadística con el fin de estimar la concordancia entre las fuentes de datos analizadas. Para recabar estos datos, se realizó un censo poblacional en la unidad de análisis y, adicionalmente, se realizaron muestras intencionadas para llevar a cabo actividades complementarias de medición. Al término del estudio, se encontraron adecuados valores de consistencia interna y concordancia entre las fuentes de medición del desempeño docente, con lo que se pudo validar el modelo revisado.

Palabras clave

Competencia didáctica, concordancia estadística, correlación estadística, evaluación, validez interna

Abstract

The objective was to validate a set of actions, which evaluates teaching competencies in a public normal school; a quantitative method was used with the application of statistical inference formulas in order to estimate the concordance between the analyzed data sources. To collect these data, a population census was performed in the unit of analysis and, additionally, purposive samples were taken to carry out complementary measurement activities. At the end of the study, adequate values of internal consistency and concordance were found among the teacher performance measurement sources, which made it possible to validate the revised model.

Keyword

Didactic competencies, statistical concordance, statistical correlation, evaluation, internal validity

Introducción

En una escuela normal pública, se identificó que los resultados de la evaluación institucional docente que se hacía de manera invariable cada semestre presentaban datos inconsistentes con la percepción y experiencia de la comunidad académica, por lo que tales datos y sus conclusiones derivadas eran cuestionadas de manera argumentada por docentes y estudiantes, dado que no mostraban correspondencia con los hechos reconocidos en la práctica diaria en las aulas.

Ante esta situación, la coordinación académica de la escuela propuso revisar el proceso de evaluación llevado a cabo, que al momento consistía en una actividad principal: una encuesta dirigida a estudiantes, realizada mediante un cuestionario en la aplicación Formularios de Google, con el que se evaluaba semestralmente a los docentes.

Mediante entrevistas a algunos estudiantes y profesores, incluyendo la directora de la institución, y el análisis de variabilidad de cinco generaciones de datos acumulados, se encontró que los resultados reportados no eran consistentes entre cohortes, además, presentaban una considerable cantidad de datos perdidos, valores extremos y, principalmente, no tenían correspondencia entre los dictámenes emitidos y los hechos conocidos sobre el desempeño de los profesores: incluso se encontraron evidentes discrepancias entre los datos numéricos recabados en la encuesta y las conclusiones correspondientes, lo que denotaba falta de validez interna en los resultados de la actividad; esta situación fue relevante y generó interés para los directivos de la escuela, ya que la intención ulterior de la evaluación docente es tomar decisiones para la mejora del funcionamiento de la docencia, lo cual resulta inviable o ineficiente, si los resultados generados de esta actividad evaluadora presentan errores aleatorios o sistemáticos de medida.

Adicionalmente, se reflexionó sobre el efecto de esta situación en el ánimo y la cultura evaluativa de la comunidad escolar. En tanto que se conocen estudios que sugieren que la atribución subjetiva de confianza que los usuarios depositan en un instrumento de medición puede tener efectos en la validez concurrente o predictiva de los datos recabados, ya que quienes responden pueden cambiar su compromiso, atención y cuidado en sus respuestas, según si consideran que el instrumento se encuentra preparado en forma adecuada o lógica, y si consideran que sus resultados son merecedores de confianza, lo que consecuentemente puede afectar de manera positiva o negativa la precisión de estas respuestas proporcionadas (Farmer, McGill & Dombrowsky, 2020).

En esta indagatoria, se encontraron también cuestionamientos sobre la pertinencia en la planificación y administración del cuestionario, y sobre el adecuado análisis de los datos recabados; lo anterior también se pensó como un aspecto sumamente importante a revisar, ya que se considera que dos condiciones básicas para otorgar confianza sobre la utilidad de las mediciones con instrumentos son, además del lógico aseguramiento

de los atributos psicométricos del material o instrumento: a) la capacidad técnica de quienes lo administran e interpretan y b) su cuidadosa administración (Muñiz, Hernández & Fernández-Hermida, 2020).

A partir de estas reflexiones, se concluyó que la acción realizada para evaluar el desempeño docente en la escuela era insuficiente para aportar datos objetivos y confiables respecto de su objeto de medida; por tanto, se planteó el objetivo de validar un modelo institucional de evaluación de la competencia docente didáctica.

Revisión de la literatura

La evaluación docente es una actividad de principal importancia en los centros y en los sistemas escolares; su práctica sistemática, reflexiva, constructiva y permanente, puede generar no solamente información objetiva para la mejora de diversas áreas de desarrollo profesional, sino que también favorece la instalación de una cultura de valoración de esta actividad, que es inherente de la propia práctica docente; es así que en todos los niveles educativos la evaluación de la docencia se ha considerado un eje para garantizar el aprendizaje de los estudiantes, ya que esta actividad evaluadora permite identificar fortalezas que conviene conservar y áreas de oportunidad que es necesario atender para superar.

Sin embargo, la materialización de este reconocimiento en la práctica representa varias dificultades, al menos en la educación superior: por principio, la evaluación del desempeño parece generar una reacción de resistencia en la mayoría de los docentes (de Diego & Rueda, 2013), a pesar de que la evaluación supone ser una práctica inherente de su labor cotidiana, y a pesar de que no pocas veces esos mismos docentes resistentes a ser evaluados, se permiten discursos exhortativos a los estudiantes para que valoren las bondades de esta actividad y miren su potencial constructivo. A razón de esto, las administraciones escolares deben realizar verdaderos esfuerzos para instalar esta práctica en las instituciones, y en el plano oficial (estatal o federal) se debe ofrecer una transacción material que motive-al menos extrínsecamente-el interés de los profesores para evaluarse de manera voluntaria.

Esta resistencia se ha explicado por la falta de claridad de las políticas educativas que promueven la evaluación docente, por las creencias conscientes y justificadas de los docentes respecto del interés puramente administrativo de las mediciones, y por la insuficiente organización o por la deficiencia de acciones y materiales implementados, lo que en conjunto generan un recelo ciertamente “natural” sobre los méritos de participar en esta revisión de su desempeño (Gómez & Valdéz, 2019; Iglesias & Loredó, 2017).

Otra dificultad que se presenta en la evaluación de los docentes radica en el carácter técnico que guarda esta actividad: sin duda, la planificación, la implementación, la valoración y el seguimiento de un sistema de evaluación demandan conocimientos y habilidades específicos en materia de la métrica cuantitativa y cualitativa de esta variable compleja, por lo que

no puede encargarse esta delicada tarea a personal sin la capacitación, recursos materiales y de tiempo de dedicación para invertir en sus diversas acciones.

Es así que se ha encontrado que en la mayoría de los casos los programas de evaluación docente se concentran en la encuesta a los estudiantes como principal o única actividad, sin que se presenten reportes de la adecuación técnica del diseño del instrumento empleado o de las actividades para su administración; por tanto, no es común encontrar el uso de más de un instrumento o técnica para el acopio de datos (Asun & Zúñiga, 2017).

Lo anterior reduce claramente la riqueza y solidez de los datos recabados, y consecuentemente, los exponen a errores sistemáticos y aleatorios de medida, o imprecisiones en el acopio de datos, que generan correspondientes imprecisiones en los resultados y las conclusiones emanados de su análisis.

En este tema específico, además, se han planteado cuestionamientos sobre la conveniencia o pertinencia de recurrir a autorreportes para recolectar información, ya que esta modalidad de informe se encuentra expuesta a varias fuentes de sesgo subjetivo o error de medida (Del Valle & Zamora, 2021; García, 2000), entre las que se pueden destacar el desinterés por la actividad, cuando no se identifica con convencimiento, así como la utilidad del esfuerzo por realizar, lo que lleva a emitir respuestas apresuradas, sin compromiso y sin detenimiento.

Por otra parte, también puede surgir incertidumbre y desconfianza sobre las implicaciones de sincerarse sobre el desempeño de los profesores, lo que también puede conllevar a la tergiversación deliberada de opiniones, a fin de mantener una posición de seguridad ante el desconocimiento de las consecuencias del acto realizado.

La situación se complica más cuando se conoce que los estudiantes pueden emplear criterios personales -y esto significa: alejados de los establecidos por las instituciones-, para evaluar la eficacia o ineficacia de sus profesores; al respecto, se puede revisar el estudio realizado por García y Medécigo (2014) con estudiantes universitarios, en el que se encontró que para esta población son más relevantes o preciados el interés y entusiasmo de los docentes por su asignatura y sus estudiantes, que su pericia o dominio disciplinar, o aun los productos o procesos de aprendizaje que se alcanzan en las clases.

Lo anterior presenta un reto teórico a resolver, por cuanto se ofrece una disonancia entre los “intereses” de las instituciones y los estudiantes respecto de lo que se considera importante en el desempeño de los docentes, y que debe resolverse para establecer cuáles son los objetivos y, por tanto, los criterios válidos para realizar la evaluación que se pretende.

Más aún, el propio objeto de medida (esto es, la evaluación del desempeño docente) representa una fuente de estudio retardadora por sí misma, ya que no existe un pleno acuerdo sobre sus delimitaciones conceptuales.

Sin embargo, existe una tendencia a considerar que la evaluación del desempeño docente debe orientarse al menos a dos grandes áreas: por una parte, a las funciones sustantivas que se reconocen en las instituciones de educación superior, como son la tutoría, investigación y otras participaciones del personal en la vida institucional; por otra parte, se encuentra el área del proceso de enseñanza y aprendizaje (Murillo & Rodríguez, 2016), que es la actividad con la que se relaciona primero a los profesores, desde la perspectiva de los estudiantes y la opinión pública, e incluso entre los docentes y directivos. En este punto, es conveniente precisar que este trabajo se concentrará en la segunda vertiente del concepto; es decir, en la orientación didáctica del desempeño del profesorado.

A pesar de estos retos planteados, prevalece la idea de que es importante llevar a cabo esfuerzos por evaluar el desempeño de los docentes, atendiendo a los reclamos sobre su planeación, su pertinencia técnica, la orientación formativa de sus resultados y la comunicación constante de sus políticas, fines y procedimientos.

Por tanto, se considera que es importante para las instituciones educativas desarrollar modelos integrales de evaluación que se adecuen a las características y los recursos de sus contextos, y que puedan consolidarse en la comunidad escolar como una estrategia reconocida para avanzar en los fines de sus programas de desarrollo (Asun & Zúñiga, 2017; Iglesias et al., 2021).

Finalmente, se enfatiza la idea de fortalecer las prácticas de evaluación para establecer un conjunto sistemático de procedimientos que se constituyan en un modelo comprensivo y eficiente que pueda implementarse de manera cabal, evitando acciones aisladas, meramente administrativas, sin fines aplicativos y resguardadas del conocimiento de la comunidad escolar.

En definitiva-se ha dicho-, un interés secundario de implementar acciones de evaluación institucional es el asentamiento de una cultura académica que reconozca las virtudes de la evaluación externa, la coevaluación, la autoevaluación y otras formas de diagnóstico del propio desempeño para incentivar el desarrollo profesional y la mejor formación de los estudiantes.

Metodología

Se diseñó un modelo integral de evaluación de la competencia didáctica que se compuso de cuatro vías de acopio de datos: a) encuesta censal a alumnos, b) encuesta a una muestra de alumnos, c) observación directa al profesor en el aula, d) análisis del producto de aprendizaje de los cursos o las asignaturas.

La hipótesis planteada para la medición de la consistencia interna del modelo se formuló como: “se encontrará un adecuado nivel de concordancia estadística entre las métricas obtenidas por las cuatro fuentes de datos”, lo que quedó expresado con la siguiente formulación: $H_0: p = k < 0.5$, donde k refiere al nivel de concordancia entre las mediciones, y el p valor se establece a nivel de 0.5, conforme a los criterios de potencia de prueba sugeridos por Cohen (1992).

Para la encuesta censal a estudiantes, se empleó un cuestionario de 12 reactivos organizados en tres factores (planeación efectiva, gestión del aprendizaje y evaluación formativa) que fue validado por Cardaña (2021), asegurando sus atributos técnicos: validez racional, validez de contenido, confiabilidad, validez criterial entre dos jueces mediante una prueba Kappa de Cohen, validez de constructo mediante un análisis factorial exploratorio, y otras pruebas post hoc, como la de adecuación muestral y de pertinencia de la elección de procesos. Este cuestionario fue aplicado mediante un formulario de Google, explicando previamente su contenido, finalidad, y garantizando el anonimato y uso estrictamente académico de los datos proporcionados por los alumnos.

Posteriormente, se realizó una segunda administración del instrumento, en esta ocasión, solamente a una muestra de 36 alumnos, seleccionados por criterio deliverado, en función de su desempeño académico; así, se trabajó con grupos conformados por los estudiantes más destacados de cada grado escolar, según su promedio de calificaciones, considerando que estos estudiantes tendrían una perspectiva disciplinada, objetiva y libre de sesgos personales sobre el desempeño de sus profesores. En este caso, se hizo la lectura de cada reactivo del instrumento antes de ser respondido, planteando precisiones o comentando preguntas de los estudiantes emitidas con el fin de clarificar las ideas pedagógicas implicadas en cada cuestión.

Para la observación directa, se empleó una escala estimativa de aspectos didácticos básicos que pueden ser observados en toda clase; este material fue diseñado y validado por la coordinación académica de la escuela; en este caso, los indicadores revisados se refieren a la planeación, la gestión del aprendizaje y el establecimiento de un ambiente propicio para el aprendizaje. La actividad se realizó en diferentes momentos del semestre y a cada profesor sólo se le observó en una ocasión en cada periodo.

Finalmente, se realizó el análisis del producto más elaborado o complejo de cada curso (asignatura) del programa educativo para identificar el tipo o nivel de aprendizaje promovido en la actividad de aprendizaje relacionada con dicho producto; para hacer este análisis, se empleó como referencia de logro la taxonomía de aprendizajes del sistema cognitivo, propuesta por Robert Marzano, que se constituye por los siguientes niveles o tipos de adquisiciones, mencionadas de menor a mayor complejidad: a) recuperación, b) comprensión, c) análisis, d) utilización del conocimiento (Gallardo, 2009; Marzano & Kendall, 2008).

Una vez generados los datos en estas vías de acopio, se realizaron dos análisis básicos, orientados a los siguientes objetivos:

a) Identificar la consistencia estadística entre los datos de la encuesta censal, la encuesta por muestra intencionada y la observación directa al profesor: en este análisis, se aplicó una prueba de concordancia con el coeficiente Kendall, dada la medida ordinal de los datos y en consideración a que esta prueba ha mostrado una adecuada adaptación a la variación de los tamaños muestrales comparados (Stockemer, 2019).

b) Identificar la consistencia estadística entre la encuesta por muestra intencionada y el análisis del producto de aprendizaje del curso; en este caso, se empleó la prueba Kappa de Cohen, que es usada principalmente para medidas nominales, pero también para ordinales; asimismo, a manera de exploración, se hizo una prueba de correlación entre estas dos variables mediante la prueba de Spearman, dado el tamaño reducido de las muestras y su distribución no paramétrica (Landeró & González, 2006; Siegel & Castellan, 2012).

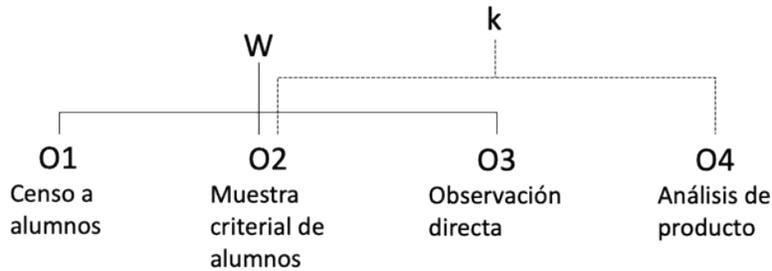
En los casos de concordancia, se emplearon los criterios de Landis y Koch (1977) para estimar la potencia de los resultados: de 0 a 0.20, consistencia leve; de 0.21 a 0.40, mediana; de 0.41 a 0.60, moderada; de 0.61 a 0.80, sustancial; de 0.81 a 1, muy fuerte o casi perfecta. Por su parte, las estimaciones de correlación se ajustaron a las convencionales propuestas por Cohen (1992), con lo que las hipótesis de prueba se establecieron formalmente como: $H_0: \rho = 0, < .5$.

Por otra parte, también se revisó la significancia estadística, estableciendo un p valor de 5%, conforme se acostumbra para las ciencias sociales (Nunnally, 1987), ajustando la prueba de hipótesis como: $H_0 > \alpha$.

Para realizar el acopio de datos, se seleccionó a un profesor común por cada uno de los cuatro grados escolares: este criterio buscaba tener un mismo desempeño a evaluar, reduciendo entonces variabilidades no explicadas por el modelo, sino por el desempeño de los sujetos evaluados; además, este profesor debía contar con al menos 5 años de servicio en la institución, para considerar que su comportamiento era estable y se explicaba por su competencia docente (habilidades, conocimientos y actitudes profesionales integradas), y no se afectaría por el proceso de adaptación a la institución.

De manera gráfica, el diseño del estudio se puede representar de la siguiente forma.

Figura 3.1
Representación gráfica del diseño metodológico



Como se observa, este modelo implica realizar tres observaciones; en este caso, de manera semestral, a las cuales se realiza dos estimaciones de concordancia, ambas no paramétricas, que se enfocan de manera especializada en una estimación de concordancia, que es el objeto de interés puesto a prueba, lo cual no impide que se puedan hacer análisis directos como la varianza típica o, mediante el ajuste de las puntuaciones a medias, análisis de varianza unifactorial.

Resultados

Al analizar los datos recabados, se obtuvieron los siguientes resultados.

Concordancia entre la encuesta censal, encuesta muestral y observación directa

Al aplicar la prueba W de Kendall, se encontró una significancia estadística igual a 0.005, lo cual permite rechazar la hipótesis nula, y retener la hipótesis de que existió concordancia entre las observaciones. Lo anterior se muestra en el reporte del procesador estadístico.

Tabla 3.1
Resultado de la prueba de concordancia Kendall

N	3
W de Kendall	0.678
Chi cuadrado	12.732
Sig. asintótica	0.005

Por otra parte, el valor de potencia de la W de Kendall (0.678) se considera un nivel sustancial de concordancia, de acuerdo con los parámetros establecidos (Landis & Koch, 1977).

Concordancia entre la encuesta muestral y el análisis de producto

De la prueba de Kappa empleada para la confrontación de los dos elementos referidos (encuesta muestral y análisis de producto), una vez que se ajustó la primera variable a la misma escala de la segunda (con cuatro categorías de desempeño), se encontraron los siguientes resultados.

Tabla 3.2
Resultado de la prueba de concordancia Kappa

Medidas simétricas				
Medida de acuerdo Kappa	Valor	Error típ. asint.	T aproximada	Sig. aproximada
	0.610	0.062	6,103	0.000

Según lo reportado, se encontró una medida de acuerdo entre las fuentes de medición considerada buena, y la prueba de significancia permite establecer una baja probabilidad de que los resultados se expliquen por un sesgo aleatorio, eliminando el tipo de error I (González, Escoto & Chávez, 2017).

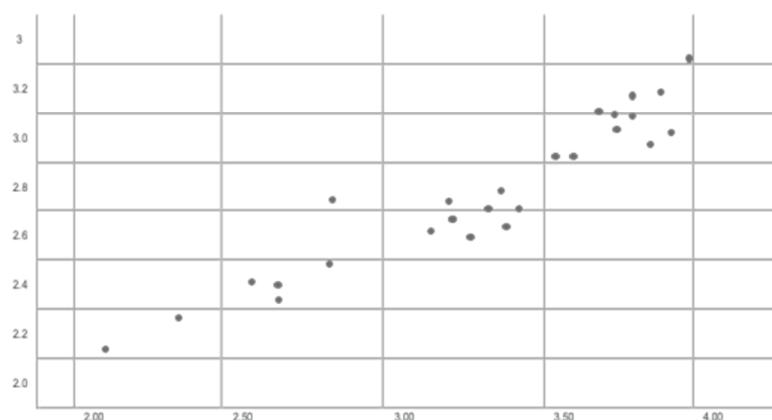
Correlación entre la encuesta muestral y el análisis de producto

Finalmente, la revisión de la correlación entre los datos mediante la prueba no paramétrica de correlación de Spearman ofreció un coeficiente de 0.661, lo cual se considera un adecuado nivel de relación entre los datos (Kerlinger & Lee, 2002), interpretándose esto como un acuerdo entre las mediciones de las fuentes.

Por otra parte, también se revisó el informe de la significancia estadística, encontrándose un p valor de 0.000, con lo que también se pudo rechazar la hipótesis nula, referida a potenciales interferencias del azar en las mediciones.

A continuación, se presenta la gráfica de puntos o nubes correspondiente a esta prueba.

Figura 3.2
Gráfica de dispersión de la prueba de correlación de Spearman



Como se puede apreciar, se presenta una dispersión lineal positiva entre las variables confrontadas, lo cual se ha propuesto como una reiteración de la concordancia entre las medidas revisadas.

Discusión

Al analizar los datos involucrados en este estudio, se encontró coincidencia con diversas aseveraciones teóricas.

La primera corroboración encontró que la evaluación a partir de la opinión de los estudiantes es el procedimiento más empleado por las instituciones educativas para el seguimiento del desempeño de sus profesores; asimismo, se reconoció el mérito o la exigencia técnica de esta actividad, al identificar diversas imprecisiones métricas en los resultados generados de esta actividad en la institución de estudio, con lo que se concluyó que, además de revisar el procedimiento existente, también merecía ser reforzado con otras estrategias o fuentes de información para considerar que sus resultados sean viables y gocen de validez interna (Universidad Complutense de Madrid, s. f.).

Lo anterior incrementa sustancialmente la complejidad de la actividad de evaluación, ya que se requería establecer mecanismos y herramientas en constante supervisión, administradas con pertinencia técnica para poder eliminar los retos de la métrica de distintos valores (medidas de las variables) y sortear diferentes efectos subjetivos humanos involucrados en los participantes, desde los propios docentes sujetos a examinación como de quienes ofrecen comentarios a su desempeño (Farmer, McGill & Dombrowsky, 2020).

Finalmente, los datos recabados del pilotaje del modelo propuesto mostraron adecuados valores de concordancia (Cohen, 1992; Siegel & Castellan, 2012), y fueron satisfactoriamente aceptados por una muestra de estudiantes evaluadores y una muestra de docentes evaluados, cuando se les comunicaron los resultados de manera individual, lo cual constituye una condición fundamental para sostener la credibilidad y el compromiso en la actividad y, consecuentemente, reducir el recelo sobre su veracidad, eficiencia y utilidad (Gómez & Valdéz, 2019; Iglesias & Loredo, 2017; Universidad Veracruzana, 2011).

Conclusiones

El modelo de evaluación docente propuesto es eficaz para generar datos válidos sobre su objeto de medida, en tanto que se encontró concordancia entre diversas fuentes de acopio de datos con métricas controladas para su corroboración.

Para la pertinente aplicación de las actividades integrantes del modelo, es importante contar con personal suficiente y capacitado para realizar las tareas básicas que implican el acopio y análisis de los datos; estas actividades se pueden mantener de manera muestral; es decir, sin necesidad de aplicarse a todos los docentes de la institución, en caso de no contar con el personal suficiente para tal fin; en cambio, se pueden realizar muestreos sobre casos específicos de interés, determinados de manera ética por los administradores académicos, con sustento en historiales u otros registros o, en su defecto, se pueden realizar selecciones aleatorias.

En la implementación del modelo, es conveniente realizar mediciones adicionales para revisar la idoneidad de las pruebas de hipótesis (González, Escoto & Chávez, 2017), por ejemplo, pruebas de igualdad de varianzas (homocedasticidad) entre los grupos de datos a comparar para evitar errores de medida que afecten las conclusiones.

Es importante realizar una permanente concientización a los estudiantes y profesores de la institución, respecto del valor de las actividades de evaluación, que deben ser comunicadas de manera oportuna, clara y con una orientación de mejora del desempeño de los docentes; asegurando la confidencialidad de las identidades de los participantes para intentar incidir en los posibles sesgos de respuesta que se pueden presentar si se desarrolla recelo ante estas tareas.

Finalmente, se propone abrir una línea de trabajo para incorporar actividades de autoevaluación y coevaluación docente, una vez que se analicen los recursos de tiempo, de personal y metodológicos disponibles, así como el estado de asimilación de las primeras actividades del modelo.

Referencias

- Asun, R. A., & Zúñiga, C. (2017). Evaluación docente universitaria: hacia una perspectiva unificada. *Revista de Sociología*, 32(1). https://www.ecorfan.org/proceedings/CDU_IX/TOMO%209_5.pdf.
- Cardeña, C. A. (2021). Validación psicométrica de un cuestionario para evaluar la competencia docente. XV Congreso de Investigación Educativa. COMIE. <https://www.comie.org.mx/congreso/2021/programa/generarProgramacionPonencias.php>.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159. <https://psycnet.apa.org/record/1992-37683-001>.
- De Diego, M., & Rueda, M. (2013). La evaluación docente en educación superior: uso de instrumentos de autoevaluación, planeación y evaluación por pares. https://www.ses.unam.mx/integrantes/uploadfile/mrueda/Rueda_DeDiego2013_LaEvaluacionDocenteEnLaEducacionSuperior.pdf.
- Del Valle, M. V., & Zamora E. V. (2021). El uso de medidas de autoinforme: ventajas y limitaciones en la investigación en psicología. *Alternativas en Psicología*, 47. <https://www.alternativas.me/attachments/article/264/El%20uso%20de%20las%20medidas%20de%20auto-informe.pdf>.
- Farmer, L. R., McGill, R. J., & Dombrowsky, S. C. (2020). Why Questionable Assessment Practices Remain Popular in School Psychology: Instructional Materials as Pedagogic Vehicles. *Canadian Journal of School Psychology (CJSP)*. <https://doi.org/10.1177/0829573520978111>.
- Gallardo, K. E. (2009). La nueva taxonomía de Marzano y Kendall: una alternativa para enriquecer el trabajo educativo desde su planeación. México: Instituto Tecnológico de Estudios Superiores de Monterrey. http://www.cca.org.mx/profesores/congreso_recursos/descargas/kathy_marzano.pdf.

- García, J. M. (2000). ¿Qué factores extraclase afectan la evaluación docente en la educación superior? *Revista Mexicana de Investigación Educativa*, 5(10), 303-325.
- García, J. M., & Medécigo, A. (2014). Los criterios que emplean los estudiantes para evaluar la ineficacia docente de sus profesores. *Perfiles Educativos*, 36(143). https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-26982014000100008.
- Gómez, L. F., & Valdéz, M. G. (2019). La evaluación del desempeño docente en educación superior. *Propósitos y Representaciones*, 7(2). <http://www.scielo.org.pe/pdf/pyr/v7n2/a19v7n2.pdf>.
- González, F., Escoto, M. del C., & Chávez, J. K. (2017). *Estadística aplicada en psicología y ciencias de la salud*. México: Manual Moderno.
- Iglesias, M., & Loredó, J. (2017). La evaluación de la docencia en educación superior: significado para el docente. XVI Congreso Nacional de Investigación Educativa. <https://www.comie.org.mx/congreso/memoriaelectronica/v14/doc/2851.pdf>.
- Iglesias, M., Loredó, J., Martínez, V., Romero, R., Alvarado, F., & Sánchez, M. (2021). Dialogremos, nuevo modelo de evaluación docente en educación superior. *Evaluación Docente*, 14(1). https://revistas.uam.es/riee/article/view/riee_14_1_001.
- Kerlinger, F. N., & Lee, H. B. (2002). *Investigación del comportamiento. Métodos de investigación en ciencias sociales*. México: McGraw Hill.
- Landero, R., & González, R. M. (2006). *Estadística con SPSS y metodología de la investigación*. México: Trillas.
- Landis, R., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1). <https://www.jstor.org/stable/i323041>.
- Marzano, R. J., & Kendall, J. S. (2008). *Designing and assessing educational objectives: Applying the new taxonomy*. Estados Unidos: Corwin Press.
- Muñiz, J., Hernández, A., & Fernández-Hermida, J. R. (2020). La utilización de los test en España: el punto de vista de los psicólogos. *Papeles del Psicólogo*, 41(1), 1-15. <https://www.papelesdelpsicologo.es/pdf/2921.pdf>.
- Murillo, M. E., & Rodríguez, M. (2016). La evaluación docente en la enseñanza universitaria. En Galván, N. (coord.), *La docencia universitaria y la formación integral de los estudiantes*. Proceedings T-IX, 44-51, México: ECORFAN.
- Nunnally, J. (1987). *Teoría psicométrica*. México: Trillas.
- Siegel, S., & Castellan, N. J. (2012). *Estadística no paramétrica aplicada a las ciencias de la conducta*. México: Trillas.
- Stockemer, D. (2019). *Quantitative Methods for the Social Sciences. A Practical Introduction with Examples in SPSS and Stata*. Estados Unidos: Springer.

Universidad Complutense de Madrid (s. f.). Modelo de evaluación de la actividad docente del profesorado de la UCM: Programa Docencia-UCM. <https://www.ucm.es/data/cont/docs/1160-2016-10-21-Modelo%20de%20evaluación%20de%20la%20actividad%20docente.pdf>.

Universidad Veracruzana (2011). Evaluación del desempeño docente. Dirección General de Desarrollo Académico. <https://www.uv.mx/psicologia/files/2013/12/Instrumento-de-evaluacion-desempeño-por-Consejo-Tecnico.pdf>.

Sobre los autores

¹ Profesor de tiempo completo de la Escuela Normal de Ticul, México, ORCID: 0000-0001-6170-2223



iQU4TRO EDITORES

En colaboración con:

