



## DISEÑO Y VALIDACIÓN DE CONTENIDO DE LA ESCALA DE EVALUACIÓN DEL DESEMPEÑO DOCENTE (EEDDOCENTE-FCAYS)

**Patricio Sebastián Henríquez Ritchie**

Universidad Autónoma de Baja California  
phenriquez@uabc.edu.mx

**Juan Carlos Pérez Morán**

Red Impulsora de Metodología en Evaluación Diagnóstica e Innovación Educativa  
jucarpint@gmail.com

**Sofía Contreras Roldán**

Investigadora

sofiacontreras@gmail.com

**Área temática:** Evaluación educativa

**Línea temática:** Diseño y validación de instrumentos

**Tipo de ponencia:** Reporte final de investigación



### Resumen

El propósito del estudio fue obtener evidencias de validez de constructo del aspecto de contenido de la Escala de Evaluación del Desempeño Docente de la Facultad de Ciencias Administrativas y Sociales (EEDDocente-FCAyS) de la Universidad Autónoma de Baja California (UABC). Dicha escala fue diseñada por la coordinación de evaluación docente de la FCAyS con el propósito de aportar información relevante sobre el desempeño docente en los programas de licenciatura de la Facultad, con base en la opinión de los estudiantes; y con ello, favorecer la toma de decisiones basada en evidencias para la mejora de la enseñanza. La validación se realizó mediante el juicio de 11 expertos en evaluación educativa y psicometría. Con base en la calificación de los expertos, se calculó la concordancia entre los jueces participantes por medio de la estimación de distintos Índices de Validez de Contenido (IVC). Como resultado, se obtuvo una V de Aiken de 0.82 ( $V \geq .69$ ); Razón de Validez de Contenido (RVC) de Lawshe de 0.82 ( $RVC \geq .56$ ); y Razón de Validez de Contenido Modificada (RVC) por Tristan de 0.91 ( $RVC' \geq .58$ ). Se concluyó, que la EEDDocente-FCAyS cumple con los criterios y estándares de validez de las pruebas establecidas a-priori. Se recomienda para futuras investigaciones ajustar los ítems señalados por los jueces y complementar con el análisis de las propiedades psicométricas mediante un estudio empírico de gran escala.

**Palabras clave:** desempeño docente, enseñanza, validez, jueceo, educación superior.

## Introducción

La evaluación docente ocupa un lugar central en las estrategias para la mejora de la calidad de la enseñanza en los sistemas educativos alrededor del mundo (Camacho, 2022; Chen y Hoshower, 2003; Horstein, 2017; Liebowitz, 2021; Salazar, 2008). En especial, el uso de escalas y cuestionarios para medir el desempeño docente desde la perspectiva de los estudiantes es uno de los mecanismos más utilizados en la actualidad por las Instituciones de Educación Superior (IES) (García, 2000; Li et al., 2018; Uttl et al., 2017; Zamora, 2021). Sin embargo, existe una preocupación creciente por el uso de dicho tipo de instrumentos de evaluación, en especial, cuando el propósito es la toma de decisiones de alto impacto (high stakes) en procesos de promoción, otorgamiento de estímulos e incentivos (Benton y Young, 2018; Boring et al., 2016; Ching, 2019; Horstein, 2017; Mitchell y Martin, 2018; Wang y Guan, 2017). Por otro lado, se reconoce la ventaja de utilizar este tipo de evaluaciones basadas en la opinión de los estudiantes para retroalimentar el desempeño y las competencias de los docentes con fines formativos (Bazán-Ramírez et al., 2021; Cashin y Downey, 1992; Li et al. 2018; Uttl et al. 2017, Silva et al., 2022).

Varios autores se han enfocado en el uso de escalas para evaluar el desempeño docente mediante la perspectiva de los estudiantes; en particular, en estudios referentes a la obtención de las propiedades psicométricas de los instrumentos de medición (Bazán-Ramírez et al., 2021; Estrada et al., 2019; Lera, et al. , 2020; Márquez y Madueño, 2016; Reyes et al., 2018). Algunos estudios a nivel internacional sobre la evaluación del desempeño docente desde la perspectiva de los estudiantes o Students' Evaluations of Teaching (SET por sus siglas en inglés) son los de Stroebe (2016), Uttl et al. (2017), Zhao et al. (2022), y García-Gómez-Heras et al. (2017). En Latinoamérica, destacan los estudios de Bazán-Ramírez y Velarde-Corrales (2021) y Márquez y Madueño (2016), en México; Bazán-Ramírez et al. (2021), en Perú; y Estrada et al. (2019), en Nicaragua.

Por su parte, Bazán-Ramírez y Velarde-Corrales (2021) realizaron un estudio en donde se evaluó el desempeño de docentes y estudiantes de una universidad peruana mediante el auto-reporte de 124 estudiantes de Psicología. La validación del instrumento utilizado en dicho estudio confirmó la estructuración teórica de cinco factores que corresponden a las cinco dimensiones de desempeño didáctico del docente (Exploración de competencias, Explicitación de criterios, Ilustración, Retroalimentación y Evaluación), derivadas de los modelos de desempeño didáctico. Los parámetros de confiabilidad (0.93) y ajuste (Chi cuadrada = 127.22;  $p = 0.00$ ; CFI = 0.90 y RMSEA = 0.07) del modelo cumplieron con los estándares y criterios de calidad que se establecen (CFI  $\geq 0.95$ , RMSEA  $\leq 0.06$ ).

Asimismo, Márquez y Madueño (2016), analizaron las propiedades psicométricas de un instrumento aplicado a estudiantes de una universidad de Sonora, con el fin de recuperar su opinión sobre las competencias básicas de los docentes en la impartición de los cursos de licenciatura. A partir de una muestra de 30,224 estudiantes, se analizó la confiabilidad y se aplicó un AFE mediante el método de componentes principales con rotación Varimax. Como resultado, se obtuvieron parámetros de confiabilidad de 0.97 y se retuvieron dos factores:

Mediación pedagógica y Actitudes docentes que explicaron el 73.1% de la varianza total de los puntajes del instrumento.

En la UABC, Henríquez et al. (2017) realizaron el primer esfuerzo de diseñar y validar una escala para evaluar el desempeño docente desde la perspectiva de los estudiantes. Dicha escala se aplicó en la Facultad de Ciencias Administrativas y Sociales (FCAYs) por primera vez en 2016, con el objetivo de evaluar dos dimensiones del desempeño docente: (a) Estrategias de enseñanza-aprendizaje y recursos didácticos y (b) Evaluación del aprendizaje. Para ello, la escala se aplicó a 738 estudiantes que evaluaron el desempeño docente de 161 profesores de los ocho programas educativos de la FCAYs: Ciencias de la educación, Administración de empresas, Derecho, Psicología, Contaduría, Sociología, Ciencias de la comunicación e Informática. Posteriormente, en 2023, se publicó el último reporte técnico del análisis de las propiedades psicométricas de la EEDDocente-FCAYs v2022-1, en la que se reportaron las evidencias de su confiabilidad, validez del aspecto de la estructura interna y de invarianza (Henríquez et al., 2023). En esta ocasión, la escala se aplicó a una muestra de 1,849 estudiantes de los programas educativos de la FCAYs. A partir de un modelo de tres factores, se realizó un Análisis Factorial Confirmatorio Multigrupo (MGCFA) para medir la invarianza de la escala por ciclo escolar y programa educativo. Como resultado, el modelo estimado obtuvo parámetros de confiabilidad ( $\alpha = 0.92$ ,  $\rho = 0.92$  y  $\omega = 0.93$ ) y ajuste del modelo adecuados ( $\chi^2 = 251.21$ ;  $gl = 87$ ;  $CFI = 0.868$ ;  $TLI = 0.841$ ;  $GFI = 0.936$ ;  $NNFI = 0.814$ ;  $RMSEA = 0.034$ ;  $SRMR = 0.057$ ). Con base en los resultados de este estudio, se puede argumentar que el modelo de tres factores de la EEDDocente-FCAYs v2022-1 puede ser utilizado para medir, desde la perspectiva de los estudiantes, el desempeño de los docentes del área de ciencias administrativas y sociales. Asimismo, la escala muestra invarianza de la medición simultánea, proporcionando evidencia para realizar análisis de diferencia de medias entre los diferentes programas educativos.

Con base en los antecedentes del estudio y evaluación del desempeño docente mediante la perspectiva de los estudiantes; se considera relevante obtener y analizar las evidencias de validez de constructo del aspecto de contenido de la nueva versión (v2023-1) de la EEDDocente-FCAYs. En especial, estimar el porcentaje de acuerdo entre jueces especialistas, y los Índices de Validez de Contenido (IVC): V de Aiken, Razón de Validez de Contenido (RVC) de Lawshe, y Razón de Validez de Contenido Modificada (RVC) por Tristan.

## Desarrollo

### *Referentes teóricos de la EEDDocente-FCAYs v2023-1*

La nueva versión de la escala (EEDDocente-FCAYs v2023-1) se ajustó con base en distintas propuestas de modelos conceptuales multidimensionales consolidados de la evaluación de la enseñanza desde la perspectiva de los estudiantes (Arreola, 2007; Bazán-Ramírez et al. 2021; Braskamp y Ory, 1994; Centra, 1993; Feldman, 1988, 1993; Fink, 2008; Marsh, 1984; 1993; 2007), y

en el análisis de sus propiedades psicométricas y adecuaciones realizadas por sus diseñadores y colaboradores (Henríquez y Arámburo, 2021; Henríquez et al., 2018; Henríquez et al., 2017; Henríquez et al., 2023). En especial, la EEDDocente-FCAYs v2023-1 presenta actualmente una estructura tridimensional para medir el desempeño docente integrada por tres factores (subescalas): (1) Organización de la clase, (2) Calidad de la enseñanza, y (3) Evaluación y retroalimentación del aprendizaje (Chan, 2018; Henríquez et al., 2023; König et al., 2017; Nasser-Abu, 2017).

### *Enfoque metodológico de la EEDDocente-FCAYs v2023-1*

La EEDDocente-FCAYs, desde sus primeras versiones, se utiliza como estrategia de evaluación docente de bajo impacto (low stakes), con un enfoque formativo y orientado hacia la mejora de los procesos de enseñanza (Bloom, 1969; Dylan, 2006; Martínez-Rizo, 2009; Popham, 2013; Scriven, 1967). Como punto central de la estrategia, se considera relevante identificar fortalezas y debilidades del desempeño de los docentes, a partir de la perspectiva de los estudiantes (Spooren et al., 2013; Zhao et al. 2022). Asimismo, se considera el uso de escalas de medición en formatos breves (un método viable, factible y adecuado) para que, con apoyo de los estudiantes, se evalúe al final de cada ciclo escolar el desempeño de los docentes que imparten clases en educación superior. Con ello, se obtiene información de forma práctica, oportuna y expedita solucionando muchos de los problemas a los que se enfrentan los evaluadores en las aplicaciones de gran escala y afectando lo menos posible las actividades escolares. Como antecedente, es importante mencionar que la FCAYs evalúa cada semestre un promedio de 330 docentes, tanto de tiempo completo como de asignatura (FCAYs, 2020), quienes imparten clases a más de 4,000 estudiantes (UABC, 2023).

Para los ajustes y adecuaciones de la EEDDocente-FCAYs v2023-1, se aplicó el modelo de medición basado en Haladyna y Rodríguez (2013), y las propuestas metodológicas del enfoque unitario de la validez de constructo (Kane, 2006; Messick, 1995). Asimismo, para su validación se siguieron las recomendaciones de los Estándares para el desarrollo de pruebas educativas y psicológicas de la American Educational Research Association, la American Psychological Association y el National Council on Measurement Education (AERA, APA y NCME, 2018), así como las directrices internacionales de la International Test Commission (ITC, 2013). En especial, para la fase de validación con jueces, se consideraron los estándares y criterios de calidad de la AERA, APA y NCME (2018) referentes al diseño de instrumentos (Estándar 4.0 al 4.14); la obtención de evidencias de validez de constructo (Estándar 1.9 y 1.10), y en particular de las evidencias de validez del aspecto del contenido (Estándar 1.11); y el análisis de la imparcialidad de los instrumentos de medición (Estándar 3.0 al 3.5 y 3.9).

### *Validación del contenido de la EEDDocente-FCAYs v2023-1*

La validación se realizó mediante el juicio de 11 expertos en evaluación educativa y psicometría. Con base en la calificación de los expertos, se calculó la concordancia entre los jueces participantes con la estimación de distintos Índices de Validez de Contenido (IVC). Como

resultado se obtuvo un Porcentaje de Acuerdo entre Jueces (PAJ) de 0.90. Asimismo, se obtuvo una V de Aiken de 0.82; Razón de Validez de Contenido (RVC) de Lawshe de 0.82; y Razón de Validez de Contenido Modificada (RVC') por Tristan de 0.91. En especial, se considera como principal método de estimación la RVC' propuesto por Tristan-López (2008) en donde la RVC' debe ser igual o superior a 0.5823, sin importar el número de panelistas. Para calcular la RVC' se calcula la fórmula:

$$RVC'' = \frac{RVC + 1}{2}$$

## Conclusiones

A modo de conclusión, se puede decir que los resultados en lo general fueron satisfactorios considerando el cumplimiento de los criterios y estándares establecidos a-priori (PAJ  $\geq$  0.80, V de Aiken  $\geq$  .69, RVC de Lawshe  $\geq$  .56, y RVC' corregida por Tristan  $\geq$  .58) de las evidencias de validez de constructo de los aspectos del contenido de la EEDDocente-FCAyS v2023-1. Además, los análisis de los IVC a nivel de los reactivos arrojaron un resultado similar en su mayoría. Fueron pocos los ítems que se identificaron con RVC' por debajo del criterio de calidad (R9.2, R9.7 y R10.5). Estos resultados indican que, una vez atendidas las observaciones y recomendaciones de los jueces para la correcciones y ajustes de los reactivos, se puede proceder al piloteo empírico para el análisis de las propiedades psicométricas y de invarianza de la EEDDocente-FCAyS v2023-1. Los hallazgos del presente estudio representan un avance en el estudio y medición del desempeño docente para fines formativos y de bajo impacto (Bazán-Ramírez et al., 2021; Cashin y Downey, 1992; Li et al. 2018; Uttl et al., 2017; Silva et al., 2022). En especial, en el desarrollo y validación de escalas para evaluar el desempeño docente, con base en la opinión de los estudiantes, con el propósito de favorecer la toma de decisiones basada en evidencias para la mejora de la enseñanza (García, 2000; Li et al., 2018; Uttl et al., 2017; Zamora, 2021). Con la información obtenida por la EEDDocente-FCAyS v2023-1 sobre las fortalezas y debilidades en dimensiones relevantes del desempeño docente se pretende detectar necesidades de actualización y formación continua para los docentes, para orientar el desempeño y las prácticas docentes a nivel del aula. Finalmente, se recomienda para futuras investigaciones realizar un piloteo empírico para el análisis de las propiedades psicométricas y de invarianza de la EEDDocente-FCAyS v2023-1.

## Referencias

American Educational Research Association, American Psychological Association y National Council on Measurement Education [AERA, APA, NCME] (2018). Estándares para el desarrollo de pruebas psicológicas y educativas. American Educational Research Association.

- Arreola, R. A. (2007). *Developing a comprehensive faculty evaluation system: A guide to designing, building, and operating large-scale faculty evaluation systems* (3rd Ed.). Bolton, MA: Anker.
- Braskamp, L. A., and Ory, J. C. (1994). *Assessing faculty work: Enhancing individual and institutional performance*. San Francisco, CA: Jossey-Bass.
- Bazán-Ramírez, A., Pérez-Morán, J. C., and Bernal-Baldenebro, B. (2021). Criteria for Teaching Performance in Psychology: Invariance According to Age, Sex, and Academic Stage of Peruvian Students. *Frontiers in Psychology*, 12. doi: 10.3389/fpsyg.2021.764081.
- Bazán-Ramírez, A., y Velarde-Corrales, N. (2021). Auto-reporte del estudiantado en criterios de desempeño didáctico en clases de Psicología. *Journal of Behavior, Health & Social Issues*, 13 (1), 22–35. doi: 10.22201/fesi.20070780e.2021.13.1.78071.
- Benton S. L. (2018). Best Practices in the evaluation of teaching. *Best Practices in Evaluation of Teaching*, 69. <https://eric.ed.gov/?id=ED588352>.
- Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. In R. W. Tyler (Ed.), *Educational evaluation: new roles, new means: the 68th yearbook of the National Society for the Study of Education* (part II) (Vol. 68(2), pp. 26-50). Chicago, IL: University of Chicago Press.
- Boring, A., Ottoboni, K., and Stark, P.B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *Sci. Open Res.* doi: 10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1.
- Camacho, I. (2022). El desempeño docente y su implicación en la enseñanza. *Formación Estratégica*, 6 (2), 105–120. doi: <https://formacionestrategica.com/index.php/foes/article/view/91>
- Cashin, W. E., and Downey, R. G. (1992). Using global student rating items for summative evaluation. *J. Educ. Psychol.* 84, 563–572. doi: 10.1037/0022-0663.84.4.563.
- Centra, J.A. (1993). *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness*. San Francisco, CA: Jossey-Bass.
- Chan, W. M. (2018). Teaching in HIGHER education: students' perceptions of effective teaching and good teachers. *Soc. Sci. Educ. Res. Rev.* 5, 40–58.
- Chen, Y., and Hoshower, L. B. (2003). Student evaluation of teaching effectiveness: An assessment of student perception and motivation. *Assessment & evaluation in higher education*, 28 (1), 71-88. doi: 10.1080/02602930301683.
- Ching, G. (2019). A literature review on the student evaluation of teaching: an examination of the search, experience, and credence qualities of SET. *High. Educ. Eval. Dev.* 12, 63–84. doi: 10.1108/HEED-04-2018-0009.
- Dylan, W. (2006). Formative Assessment: Getting the Focus Right, *Educational Assessment*, 11:3-4, 283-289, DOI: 10.1080/10627197.2006.9652993.
- Estrada, L. A., Yglesias, L. A., Miranda, A. E., Díaz, J. K., y Díaz, S. M. (2019). Propiedades Psicométricas de un Cuestionario sobre Evaluación del Desempeño Docente Universitario desde la Percepción del Estudiante. *Revista de Investigación Estadística*, 2(1), 92-102. <https://revistas.unitru.edu.pe/index.php/REDIES/article/view/2849>.

- Facultad de Ciencias Administrativas y Sociales [FCAYs], Universidad Autónoma de Baja California (2020). Tercer informe de actividades. <https://fcays.ens.uabc.mx/wp-content/uploads/2020/11/061120-Tercer-Informe-FCAYS.pdf>
- Feldman, K. A. (1993). College students' views of male and female teachers: Part II- evidence from students' evaluations of their classroom teachers. *Research in Higher Education*, 34(2), 151-211. <https://www.jstor.org/stable/40196104>.
- Feldman, K. A. (1988). Effective college teaching from the students' and faculty's view: Matched or mismatched priorities? *Research in Higher Education*, 28(4), 291-344. doi: 10.1007/BF01006402.
- Fink, D. (2008). Evaluating teaching: A new approach to an old problem. In S. Chadwick-Blossey and D. R. Robertson (Eds.), *To improve the academy: Resources for faculty, instructional, and organizational development* (Vol. 26, pp. 3-21). Jossey-Bass.
- García, J. M. (2000). Las dimensiones de la efectividad docente, validez y confiabilidad de los cuestionarios de evaluación de la docencia: síntesis de investigación internacional. In M. Rueda y F. Díaz-Barriga (comp.), *Evaluación de la docencia* (pp. 41-62). Paidós. <https://bit.ly/2HFMvoW>.
- García-Gómez-Heras, S., Gil-Madrona, P., Hernandez-Barrera, V., Lopez-de-Andres, A., and Carrasco-Garrido, P. (2017). Importance of university teacher behaviour in the faculty of health science. *Australasian Medical Journal*, 10(9), 800–810. <https://doi.org/10.21767/AMJ.2017.3128>.
- Haladyna, T. M. and Rodríguez, M.C. (2013). *Developing and Validating Test Items*. Routledge. doi:10.4324/9780203850381.
- Henríquez, P., Pérez-Morán, J. C., Cid, C. D. y Zamora, J. E. (2023). Reporte técnico de las propiedades psicométricas de estructura factorial e invarianza de la Escala de Evaluación del Desempeño Docente (EEDDocente) de la FCAYs. Documento interno.
- Henríquez, P. y Arámburo, V. (2021). Evaluación del desempeño docente por áreas de conocimiento: El caso de la Facultad de Ciencias Administrativas y Sociales de la Universidad Autónoma de Baja California, México. *Actualidades Investigativas. Educación*, 21(3), 1–20. <https://doi.org/10.15517/aie.v21i3.46294>.
- Henríquez, P., Arámburo, V. y Boroel, B. (2018). Análisis de las estrategias de enseñanza según áreas de conocimiento en el nivel educativo superior: Percepciones de estudiantes de la Facultad de Ciencias Administrativas y Sociales (FCAYS), Universidad Autónoma de Baja California (UABC), México. *Compendio Investigativo de Academic Journals Celaya 2018*, tomo 14, 2320-2325. Disponible en: <https://www.academiajournals.com/pub-celaya-2018>.
- Henríquez, P., Arámburo, V. y Dávila, E. (2017). Percepción de los estudiantes universitarios acerca de las estrategias pedagógicas y de evaluación del aprendizaje utilizadas por sus profesores: el caso de la FCAYS de la UABC. *Memorias electrónicas del XIV Congreso Nacional de Investigación Educativa, COMIE*. San Luis Potosí, México.
- Hornstein, H. A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education*, 4(1). doi:10.1080/2331186x.2017.1304016.

- International Test Commission. (2013). ITC Guidelines on Test Use. ITC-G-TU-20131008. [https://www.intestcom.org/files/guideline\\_test\\_use.pdf](https://www.intestcom.org/files/guideline_test_use.pdf).
- Kane, M. (2006). Validation. En R. L. Brennan (Ed.), *Educational measurement* (pp. 17-64). Westport: National Council on Measurement in Education and American Council on Education.
- König, J., Ligtoet, R., Klemenz, S., and Rothland, M. (2017). Effects of opportunities to learn in teacher preparation on future teachers' general pedagogical knowledge: analyzing program characteristics and outcomes. *Stud. Educ. Eval.* 53, 122–133. doi: 10.1016/j.stueduc.2017.03.001.
- Lera, M., León, J. and Ruiz, P. (2021). Adaptation of the Teacher Efficacy Scale to Measure Effective Teachers' Educational Practices Through Students' Ratings: A Multilevel Approach. *Psicothema*, 33(3), 509-517. doi: <https://doi.org/10.7334/psicothema2020.262>
- Li, G., Hou, G., Wang, X., Yang, D., Jian, H., and Wang, W. (2018). A Multivariate Generalizability Theory Approach to College Students' Evaluation of Teaching. *Frontiers in Psychology*, 9. doi:10.3389/fpsyg.2018.01065.
- Liebowitz, D. (2021). Teacher evaluation for accountability and growth: Should policy treat them as complements or substitutes? *Labour Economics*, 71, 102024. doi: <https://doi.org/10.1016/j.labeco.2021.102024>.
- Márquez, L. y Madueño, M. L. (2016). Propiedades psicométricas de un instrumento para apoyar el proceso de evaluación del docente universitario. *Revista Electrónica de Investigación Educativa*, 18(2), 53-61. <http://redie.uabc.mx/redie/article/view/793>.
- Martínez-Rizo, F. (2009). Evaluación formativa en aula y evaluación a gran escala: hacia un sistema más equilibrado. *Revista Electrónica de Investigación Educativa*, 11(2). <http://redie.uabc.mx/vol11no2/contenido-mtzrizo2.html>.
- Marsh, H.W. (2007). Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases and Usefulness. In: Perry, R.P., Smart, J.C. (eds) *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*. Springer, Dordrecht. doi: 10.1007/1-4020-5742-3\_9.
- Marsh, H. W. (1993). Multidimensional students' evaluations of teaching effectiveness. *Journal of Higher Education*, 64(1), 1-18. doi: 10.1080/00221546.1993.11778406.
- Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76(5), 707-754. doi: 10.1037/0022-0663.76.5.707.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749. <https://doi.org/10.1037/0003-066X.50.9.741>.
- Mitchell, K., and Martin, J. (2018). Gender Bias in Student Evaluations. *PS: Political Science & Politics*, 51(3), 648-652. doi:10.1017/S104909651800001X.

- Nasser-Abu, F. (2017). Teaching in higher education: good teaching through students' lens. *Stud. Educ. Eval.* 54, 4–12. doi: 10.1016/j.stueduc.2016.10.006.
- Popham, W. J. (2013). *Evaluating America's teachers: Mission possible?*. Corwin Press.
- Reyes, E., Luna, E. y Caso, J. (2020). Evidencias de validez del Cuestionario de Evaluación de la Competencia Docente Universitaria. *Perfiles Educativos*, 42(169), 106-122. <https://doi.org/10.22201/iisue.24486167e.2020.169.58931>.
- Salazar, J. (2008). Diagnóstico Preliminar sobre Evaluación de la Docencia Universitaria. Una Aproximación a la Realidad en las Universidades Públicas y/o Estatales de Chile. *Revista Iberoamericana de Evaluación Educativa* 1:3e. <https://revistas.uam.es/riee/article/view/4655>.
- Scriven, M. (1967). The methodology of evaluation. In *Perspectives on curriculum evaluation*, Edited by: Tyler, R, Gagne, R and Scriven, M. Chicago: Rand McNally. AERA monograph series – curriculum evaluation.
- Silva, J., Solís, P., Huaman, J. y Quispe, F. (2022). La evaluación formativa en el desempeño docente universitario: Revisión sistemática de literatura. *TecnoHumanista Revista Científica*, 2(4), 1-14. doi: 10.53673/th.v2i4.177.
- Stroebe, W. (2016). Why Good Teaching Evaluations May Reward Bad Teaching: On Grade Inflation and Other Unintended Consequences of Student Evaluations. *Perspectives on Psychological Science*, 11(6), 800–816. <https://doi.org/10.1177/1745691616650284>.
- Spooren, P., Brockx, B., and Mortelmans, D. (2013). The validity of student evaluation of teaching. *Rev. Educ. Res.* 83, 598–642. doi: 10.3102/0034654313496870.
- Tristán-López, A. (2008). Modificación al modelo de Lawshe para el dictamen de validez de contenido de un instrumento objetivo. *Avances en Medición*, 6, n. 1: 37–48.
- Universidad Autónoma De Baja California [UABC] (2023). Registro De Estadística Poblacional Periodo Escolar 2023-1. <http://cgsege.uabc.mx/documents/10845/68433/Poblaci%C3%B3n%20Estudiantil%202023-1>.
- Uttl, B., White, C. A., and González, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54, 22–42. doi:10.1016/j.stueduc.2016.08.007.
- Wang, D. F., and Guan, L. (2017). Higher education quality evaluation from the perspective of students: theoretical construction and reflection. *J. Nat. Inst. Educ. Admin.* 5:75. doi: 10.3969/j.issn.1672-4038.2017.05.005.
- Zamora, E. (2021). La evaluación del desempeño docente mediante cuestionarios en la universidad: Su legitimidad según la literatura y los requerimientos para que sea efectiva. *Revista Actualidades Investigativas en Educación*, 21(3), 1-23. doi: 10.15517/aie.v21i3.46221.
- Zhao L., Xu P, Chen Y., and Yan, S. (2022). A literature review of the research on students' evaluation of teaching in higher education. *Frontiers in Psychology*, 13:1004487. doi:10.3389/fpsyg.2022.1004487.