

Introducción

- Este documento presenta el análisis de evidencias de validez de las puntuaciones de un Instrumento de Evaluación de la Docencia [IED] de una universidad chilena. Para para ello se analizó la fiabilidad del instrumento y se calcularon correlaciones de Pearson entre las puntuaciones del IED y el rendimiento académico y las experiencias de aprendizaje de los estudiantes. Además, se relacionaron las respuestas abiertas de los estudiantes con las puntuaciones asignadas. Finalmente, se desarrolló un cuestionario para evaluar las consecuencias de los resultados del Instrumento.
- La evidencia a favor de la validez de las puntuaciones del IED implica alta confiabilidad; relación concurrente y moderada entre el IED y las experiencias de aprendizaje de los estudiantes; y los usos descriptivos y formativos de los resultados por parte de los docentes. Características menos auspiciosas del instrumento son su unidimensionalidad, su débil asociación con el rendimiento académico de los estudiantes, y su imposibilidad para discriminar las percepciones de los estudiantes sobre la calidad de la docencia.
- Estos resultados aportan antecedentes relevantes para revisar los procesos de evaluación de la docencia en otras universidades y promover la mejora de las prácticas educativas.

Sobre la recopilación de evidencias de validez

La validez se comprende como la condición de que un test presente evidencias de estar midiendo lo que dice estar midiendo, respecto de la finalidad para la cual fue construido (Hogan, 2004). En ese sentido, la validez de un instrumento no se puede reducir a un índice o a medidas de confiabilidad

(Kane, 2013), sino que se trata de una serie de evidencias que avalan (o no) su grado de validez respecto de los fines esperados.

El Standards (American Educational Research Association [AERA], American Psychological Association [APA], National Council on Measurement in Education [NCME], 2014) entiende que la evidencia que se acumula sobre la validez de un test se puede organizar en cinco tipos: evidencia basada en el contenido del test; evidencia basada en los procesos de respuesta; evidencia basada en la estructura interna; evidencia basada en relaciones con otras variables; y evidencias basadas en las consecuencias de la medición (AERA, et al., 2014).

Asimismo, Kane (2013) señala que para validar el uso de los resultados de los tests es necesario evaluar sus cuestionamientos como plausibles. La validación es entonces una evaluación de la coherencia y la integridad de la interpretación y uso de los razonamientos sobre las puntuaciones de un instrumento, y de la plausibilidad de sus inferencias y suposiciones. En ese sentido, Kane (2001) propone diferenciar entre la “parte descriptiva” y “prescriptiva” del argumento interpretativo, donde la primera implica la estimación de la variable observada en los sujetos, mientras que lo prescriptivo destaca las interpretaciones para la toma de decisiones sobre las puntuaciones obtenidas.

En este estudio, la validez se comprenderá como la construcción de un argumento basado en evidencias que respaldan las interpretaciones de las puntuaciones de los instrumentos. Estos argumentos pueden referir a sus consecuencias esperadas y no esperadas, descriptivas y prescriptivas, y derivadas del uso del instrumento en la comunidad universitaria. La investigación plasmada en este documento presenta diversas evidencias de validez sobre el IED de una universidad chilena, con el fin de proyectar orientaciones sobre su diseño y uso en otros contextos educativos.

Evidencias de validez y evaluación de la docencia universitaria: una revisión de literatura

Confiabilidad de los IED

Las investigaciones reportan altos niveles de consistencia interna de los IED, utilizando métodos de Bipartición y Alpha de Cronbach, cuyos resultados varían entre 0.77 a 0.94 (Costin, et al. 1971).

Evidencias sobre el contenido

Al analizar las evidencias de validez respecto del contenido de los IED, la primera preocupación es conceptualizar la docencia de calidad. Tanto Coll y Solé (2001) como Shulman (1989), coinciden en que el comportamiento observable del profesor es la variable más significativa en el rendimiento académico de los estudiantes. Esta perspectiva generó los populares “modelos de enseñanza eficaz” (Zabalza, 2009).

Sin embargo, criticando esta perspectiva por simple y unidimensional, Shulman (1989) la denominó “*paradigma proceso – producto*”, ya que otras variables del proceso de enseñanza y aprendizaje – como el contexto de aula, contenidos de la enseñanza y motivaciones de los estudiantes – resultan prácticamente ausentes. En respuesta a esta crítica, las investigaciones comenzaron a considerar el contexto de aula y las relaciones pedagógicas en el aprendizaje para caracterizar una enseñanza de calidad (Biggs, 2008; Coll & Sánchez, 2008; Coll y Solé, 2001). Según el autor, una buena enseñanza puede seguir distintos métodos, pero se centra en la preocupación por las necesidades del estudiante, utiliza métodos de evaluación apropiados para promover el aprendizaje profundo y procura definir metas y objetivos claros (Ramsden, 2003).

Esta perspectiva teórica ha dado paso a una serie de instrumentos que se han centrado en reconocer los distintos factores que componen la interacción en el aula. Entre estos destaca el Course Experience Questionnaire (Webster, Chan, Prosser & Watkins, 2009), que mide la experiencia de aprendizaje de los estudiantes en un curso o programa en particular, a través de las escalas “buena docencia”; “metas y objetivos claros”; “carga de trabajo apropiada” y “evaluación apropiada” (Webster, et al., 2009).

Sin embargo, la dificultad principal con los instrumentos de evaluación de la docencia está en que muchos administradores universitarios no siguen ninguna de éstas perspectivas teóricas de manera clara en el desarrollo de los instrumentos (Ory & Ryan, 2001; Penny, 2003). Por ello se

destaca que, a pesar de que los IED normalmente gozan de validez aparente (Onwuegbuzie et al., 2009), existen grandes diferencias entre como definen una “buena docencia universitaria” (Richardson, 2005).

Evidencias sobre la estructura interna

La investigación también ha aportado evidencias sobre la relación entre las dimensiones de calidad docente que los estudiantes reconocen respecto a aquello que aparece en los instrumentos. Onwuegbuzie et al. (2007) señalan que los estudiantes reconocen como categorías asociadas a la enseñanza efectiva la docencia centrada en el estudiante, la experticia docente, el profesionalismo del profesor, su entusiasmo, su comportamiento ético, la retroalimentación efectiva, la capacidad comunicativa del profesor, la accesibilidad del equipo docente, la capacidad de liderazgo del docente, y el ambiente de aprendizaje efectivo. Estos autores también concluyen que las cuatro primeras características son las más relevantes y determinantes para los estudiantes, mientras que las categorías “docencia centrada en el estudiante”, “experticia”, “entusiasmo” y “comportamiento ético” no están presentes en el cuestionario de evaluación docente que ellos responden en su universidad. Es decir, tres de las cuatro características más relevantes para los estudiantes sobre la docencia efectiva, no están considerados en este instrumento de evaluación de la docencia.

Otras investigaciones (Ambady & Rosenthal, 1993; Shevlin et al., 2000) concluyen que los estudiantes basan sus respuestas en una sola característica del docente y no consideran otros aspectos incluidos en los cuestionarios; en otras palabras, los estudiantes estarían respondiendo bajo un “efecto halo” que los lleva a responder de una determinada forma las evaluaciones.

Respecto de evidencias sobre la estructura interna de los instrumentos, variadas investigaciones revelan que, a pesar de que normalmente los instrumentos consideran de 3 a 5 dimensiones en sus diseños, los IED se caracterizan por su unidimensionalidad (Richardson, 2005; García, 2000). Ese constructo general ha sido denominado “habilidad general para enseñar” (Abrami, et al., 1991). Spooren et al., (2013) argumentan que dichos antecedentes sobre la unidimensionalidad de los IED, deben constituirse en una preocupación para las autoridades universitarias, ya que una estructura unidimensional podría ser apropiada para fines los sumativos de los IED – en términos de facilitar el reporte de un puntaje total – pero no para sus fines formativos –

por ser más restrictivos en cuanto al tipo de información reportada a los docentes.

Evidencias sobre relación con otros criterios de calidad de la docencia

Otras investigaciones (Clayson, 2009) han centrado su quehacer en establecer relaciones entre las puntuaciones de los IED y otros criterios asociados a la práctica docente, incluyendo los resultados de aprendizaje de los estudiantes. Algunos estudios muestran correlaciones positivas entre las calificaciones de los estudiantes y las puntuaciones IED de sus profesores que varían entre 0,10 y 0,47 (Cohen, 1981; Griffin, et al., 2014); este último estudio encuentra correlaciones positivas y moderadas ($r=0,37$ $p < 0,01$, $n=2073$) ambos indicadores (Griffin, et al., 2014).

Por otro lado, el meta análisis de Clayson (2009) reporta que la percepción que los estudiantes tienen sobre su aprendizaje se relaciona con los puntajes en las evaluaciones docentes que asignan a sus profesores, aunque estas relaciones son débiles y varían de acuerdo a la disciplina y a la forma en que se mide el aprendizaje (Clayson, 2009). Por ejemplo, las relaciones son más fuertes en el área de la educación y las artes que en las áreas de negocios; y mientras más “objetiva” es la medición del aprendizaje menos fuerte es la relación entre los IED y el rendimiento de los estudiantes (Clayson, 2009).

Evidencias sobre las consecuencias del uso de los IED

Las consecuencias de los IED se pueden reconocer en términos descriptivos (orientados al diagnóstico) y prescriptivos (orientados al cambio). Entre las *consecuencias descriptivas* la literatura señala que las puntuaciones de las evaluaciones docentes son utilizadas para diferenciar a los profesores en categorías de calidad (Beran, et al., 2005) y reconocer los niveles de satisfacción de los estudiantes con sus experiencias de aprendizaje (Huybers, 2013). Respecto de las *consecuencias prescriptivas*, las puntuaciones de los IED deberían ser utilizadas con fines formativos de modo de aportar elementos básicos para la autocrítica docente (Salazar, 2008; Huybers, 2013); ofrecer información para mejorar la calidad de la docencia (Huybers, 2013; Salazar, 2008; Spooren et al. 2013); y promover que los docentes busquen instancias de apoyo en formación pedagógica (Nasser y Fresko, 2001). Sin embargo, se ha documentado que los resultados se han utilizado

con fines divergentes a los sugeridos, como por ejemplo, para orientar la permanencia, promoción e incentivos de los docentes universitarios (Huybers, 2013); y para reportar la calidad global de la enseñanza de las carreras en procesos de acreditación, proyectos, entre otros (Moore y Kuol, 2005). Cabe señalar que las investigaciones en dicha línea, coinciden en que el uso de los instrumentos con fines sumativos (de promoción, retención e incentivos) trajo consigo que sus propósitos formativos (mejora de la calidad docente) se distorsionaran (Spooren et al, 2013).

A esta variedad de posibles decisiones o acciones derivadas del uso de las encuestas de evaluación docente, se suman las diferencias entre los enfoques con los que los agentes, legítimamente implicados en el uso de los IED (administradores, decanos, docentes, estudiantes, etc.), valoran el proceso de evaluación juzgando las consecuencias desde diferentes perspectivas, intereses y valores (Padilla, Gómez, Hidalgo y Muñiz, 2007).

Un estudio conducido por Beran et al., (2005) analizó los usos y consecuencias de los IED en estudiantes¹, docentes y directores, a partir de la aplicación de encuestas a muestras representativas de cada uno de los actores señalados. Entre sus principales conclusiones es posible advertir que un 52% de los estudiantes declaran que nunca han utilizado los resultados de los IED, mientras que el resto de ellos indica que utiliza los resultados para contar con información adicional al momento de seleccionar asignaturas. Por otra parte, el estudio revela opiniones favorables de los profesores, respecto al uso de los IED como antecedente que aporta a la mejora de la docencia, sin embargo, paradójicamente ellos señalan que generalmente no los utilizan para incorporar cambios a su enseñanza. Finalmente, los directivos o administradores coinciden en que utilizan los IED para tomar decisiones relativas a los incentivos o retención de sus docentes, más no con otros fines (Beran et al. 2005).

Metodología de recopilación de evidencias de la presente investigación

Desde un método no experimental, transversal de alcance descriptivo, esta investigación cuenta con una metodología de carácter mixta subsecuente, es decir, se abordará de manera cualitativa y cuantitativa, a partir de los resultados

¹ Cabe señalar que el estudio fue realizado en Canadá, donde los resultados de las evaluaciones docentes son entregados a los estudiantes.

del IED de una universidad chilena, pública y selectiva.

Confiabilidad del instrumento

La confiabilidad del instrumento fue medida a partir de su consistencia interna, a través de Alfa de Cronbach y el método de bipartición.

Estructura interna del instrumento de evaluación de la docencia

Se aplicó un Análisis Factorial Exploratorio (AFE) y Análisis Factorial Confirmatorio (AFC) a una muestra de 48.129 respuestas de cerca de 2.000 estudiantes de primero a quinto año a la evaluación de la docencia del primer semestre del año 2014.

Relación entre el IED y el rendimiento académico de los estudiantes

A través de un análisis de correlación de Pearson, se determinó la relación entre las puntuaciones IED y las calificaciones estandarizadas de los estudiantes en cada una de las asignaturas evaluadas.

positivo en el desarrollo de la asignatura” y “señale brevemente algún aspecto que podría mejorar el desarrollo de la asignatura”. La codificación fue deductiva de tipo temática (Braun & Clarke, 2008). El análisis fue realizado por dos investigadores independientes, que codificaron de manera cruzada un 20% de los comentarios de cada carrera. Posteriormente, los códigos fueron agrupados en su frecuencia de aparición para cada uno de los nivel de logro (alto, medio y bajo) y comparados respecto de su nivel de representación entre sí.

Consecuencias del IED

Se diseñaron dos encuestas, una para docentes y una para directivos, sobre las consecuencias descriptivas y prescriptivas, esperadas y no esperadas, declaradas en la literatura sobre este tipo de instrumentos (Huybers, 2013; Moore y Kuol, 2005; Nasser y Fresko, 2001; Spooren et al. 2013; Onwuegbuzie et. al., 2007; Richardson, 2005). 204 docentes (10% del total) y 31 directivos (33% del total) contestaron las encuestas.

Tabla N°1.

Análisis Factorial Confirmatorio Encuesta Evaluación de la Docencia

		IED (1escala)	IED (4 escalas)			
			1	2	3	4
Número de ítem		21	3	9	6	3
Coeficiente de confiabilidad		.9786	.9443	.9551	.9386	.9280
Modelo con covarianzas restringidas	RMSEA < 0.08	.066		0.092		
	CFI > 0.9	.977		0.952		
	TLI > 0.9	.972		0.944		
	SRMR < 0.05	.015		0.026		

Fuente. Elaboración propia

Relación entre el IED y percepción sobre la experiencia de aprendizaje de los estudiantes

Se analizó la relación entre las puntuaciones de la encuesta de evaluación de la docencia y las puntuaciones de 311 estudiantes (de 19 cursos distintos de 5 áreas de conocimiento) en el instrumento Course Experience Questionnaire (CEQ). En particular, se calcularon correlaciones de Pearson entre las puntuaciones del CEQ y las puntuaciones del IED.

Relación entre el IED y percepción de los estudiantes sobre la calidad de la docencia

El análisis se hizo a partir de la codificación de 5991 respuestas abiertas de los estudiantes a 826 docentes de las siguientes preguntas del IED: “señale brevemente algún aspecto que considere

Resultados

Se evidencia alta **consistencia interna** del instrumento IED, con una fiabilidad de .978 en Alfa de Cronbach y .924 en Bipartición. En el análisis de estructura interna del instrumento, el AFE arrojó sólo un autovalor mayor que 1, el cual explica el 70% de la varianza, lo que coincide de manera evidente con el gráfico de sedimentación. Por otra parte, el AFC (ver Tabla N°1) demuestra que los índices de ajuste están en los parámetros esperados para el modelo unidimensional, más no para el modelo multidimensional. Esto no permite aceptar el modelo de cuatro dimensiones

Por otra parte, las puntuaciones del IED se **relacionan de manera positiva pero débil** con el rendimiento académico de los estudiantes que

responden este instrumento ($r = 0,214$; $p < 0,01$; $n=2.237$). Cabe mencionar que la intensidad de las correlaciones varía entre las áreas de conocimiento. Las correlaciones más robustas (aunque siempre moderadas) se manifiestan en las carreras de humanidades ($r = 0,32$; $p < 0,01$; $n=549$) y en las ciencias de la salud ($r = 0,43$; $p < 0,01$; $n=48$). Mientras que en las carreras del área de las ingenierías, ciencias matemáticas y ciencias naturales, las correlaciones son muy bajas (ingeniería: $r = 0,1$; $p < 0,01$; $n=688$; ciencias matemáticas: $r = 0,1$; $p < 0,01$; $n=247$) e incluso no significativas en términos de significancia estadística (ciencias naturales: $r = 0,1$; $p > 0,01$; $n=122$).

respuestas a la pregunta “elementos que debería mantener” demuestran que no existen diferencias notorias entre los docentes con puntuaciones altas, puntuaciones medias y puntuaciones bajas (estos últimos fueron muy escasos para el análisis y por eso no se consideran en la tabla), salvo en el código “preocupado por necesidades de estudiantes” (diferencia de 3,2%). Cabe mencionar que respecto a la pregunta por “sugerencias de mejora” en ambos grupos son frecuentes los comentarios asociados a “métodos y estrategias inapropiadas” y “métodos de evaluación inválidos”. Es decir, elementos de la dimensión “profesionalismo docente”. Cabe destacar que no existen diferencias determinantes por áreas de conocimiento.

Tabla N°2.

Correlación entre evaluación de la docencia y experiencia de aprendizaje de los estudiantes

		Correlaciones			
		Buena docencia	Metas y Objetivos claros	Evaluación apropiada	Carga de trabajo apropiada
Evaluación de la docencia	Correlación de Pearson	,501***	,451***	,234***	,314***
	Sig. (bilateral)	,000	,000	,000	,000
	N	311	311	311	311

Fuente. Elaboración propia. *** $p < 0,01$; ** $p < 0,05$; * $p < 0,1$

Tabla N°3.

Relación evaluación de la docencia y percepción de la calidad de la docencia (respuestas a pregunta: elementos que el docente debería mantener)

	Agrupación promedios de los IED de docentes					
	Puntuaciones altas (4 y 5)	%	Puntuaciones medias (3,01 a 4)	%	Diferencia puntuaciones altas y medias	
Buen profesor	425	12,7	81	12,2	0,5	
Métodos y estrategias apropiadas	296	8,8	47	7,1	1,7	
Claridad para explicar	276	8,2	38	5,7	2,5	
Aprendizaje activo	267	8	37	5,6	2,4	
Preocupado por necesidades de estudiantes	243	7,3	27	4,1	3,2	
Características personales (amable, simpático)	203	6,1	33	5	1,1	
Responde claramente a las dudas	167	5	35	5,3	0,3	

Fuente. Elaboración propia

Asimismo, se evidencian **correlaciones positivas y moderadas entre las puntuaciones del IED y la experiencia de aprendizaje** de los estudiantes en todas las escalas del CEQ. Las correlaciones más altas fueron con las dimensiones “Buena Docencia” y “Metas y Objetivos Claros” (ver Tabla N°2).

Sobre la relación **entre las puntuaciones del IED con la percepción de los estudiantes sobre la calidad de la docencia** (ver Tabla N°3), las

Sobre las **consecuencias de instrumento**, más del 70% de los docentes utiliza “siempre” o “casi siempre” los resultados con fines descriptivos, para conocer la valoración sobre su desempeño. Más de un 50% declara utilizar los resultados de sus evaluaciones con fines prescriptivos formativos, para mejorar sus prácticas docentes. Mientras que más de un 40% declara utilizar “siempre” o “casi siempre” los resultados de sus evaluación con fines prescriptivos sumativos, del tipo “argumentar mi

continuidad en la institución”. Al mismo tiempo, más del 70% de los directivos declaran que “siempre” o “casi siempre” utilizan los resultados de la evaluación con fines descriptivos. Un 60% declara que “a veces” o “nunca” utilizan los resultados con fines prescriptivos formativos. Y, sobre los fines prescriptivos sumativos, la mayoría de los directivos (más de 60%) declaran utilizar los resultados con fines de reporte de datos (proyectos o acreditación) pero sólo “a veces” o “nunca” (más de 65%) con fines de despido o premiación.

Sobre sus comentarios al proceso, se destaca que tanto docentes como directivos critican la aplicación del instrumento por generar respuestas “poco objetivas” por parte de los estudiantes. Esto lo asocian a sesgos de “poca conciencia” de los estudiantes en las respuestas y a la contaminación del rendimiento académico obtenido por los estudiantes en el curso. También hay críticas sobre el contenido del instrumento, respecto a su escasa coincidencia con algunos contextos de enseñanza. Otras críticas abordan la inoportuna entrega de resultados y la necesidad de incorporar a otros actores en el proceso de evaluación.

En relación a los fines y consecuencias, algunos directivos declaran discutir formativamente los resultados con sus equipos, mientras que otros destacan la existencia de algunas condiciones institucionales que no permiten tomar decisiones de despido aunque ellos lo desearan, por ejemplo.

Implicancias para las instituciones de educación superior

Los IED demuestran alta consistencia interna (Richardson, 2005), es decir, demuestran alta estabilidad en sus resultados. Sin embargo, y pese a sus objetivos, su estructura interna unidimensional no permite avanzar hacia una retroalimentación basada en dimensiones, (García, 2000; Richardson, 2005), generando menor riqueza al momento de orientar la mejora de la docencia (Spooren, et. al., 2013). Esta discrepancia entre el contenido y el uso de los IED debería alertar a las autoridades universitarias.

A lo anterior se suman los comentarios de docentes y directivos respecto a que los estudiantes responden a la encuesta con ciertos sesgos. Según la literatura, las instituciones deben tener presente el denominado “efecto halo” que implica que lo que se evalúa podría ser más bien “habilidad general para enseñar” (Abrami, et al., 1991) o, según la perspectiva de Ramsden (2003), la valoración de los estudiantes sobre su experiencia de aprendizaje. Siguiendo este último argumento y a partir de la

relación concurrente entre el IED y la experiencia de aprendizaje de los estudiantes, las instituciones de educación superior deben tener presente que los estudiantes están considerando su propia experiencia de aprendizaje en la asignatura al momento de evaluar a sus profesores, más que las dimensiones que evalúa el IED en sí.

De la misma forma, las instituciones deberían estar atentas a la diferenciación que hacen los estudiantes sobre la “docencia centrada en el estudiante” y el “profesionalismo docente” como elementos distintivos en la calidad de la docencia. Esto coincide parcialmente con el estudio de Onwuegbuzie et al., (2007) donde se menciona que lo que más valoran los estudiantes no es necesariamente lo que preguntan los cuestionarios. En ese sentido, las instituciones deberían considerar incluir en los IED las perspectivas de los estudiantes en relación a la calidad de la docencia.

Por otro lado, y a diferencia de lo esperable, los resultados de esta investigación y de otras (Clayson, 2009) muestran pequeñas correlaciones entre el IED y el rendimiento académico. Cabe mencionar que esta evidencia contradice las afirmaciones de algunos docentes y directivos, que aseguran la existencia de sesgo entre las calificaciones de los estudiantes y las puntuaciones del IED. Sin embargo, son atendibles las diferencias en las áreas de conocimiento que advierten que las disciplinas humanistas – quizás con métodos de medición vinculados a situaciones de evaluación de desempeño, como ensayos o demostraciones – se relacionan de manera más fuerte con los IED que las áreas de conocimiento más cercanas a las ciencias exactas – con métodos de evaluación asociados a pruebas de opción múltiple. Esto refuerza la importancia de adaptar los IED a distintos contextos educativos, debido a las diferencias que podría tener su utilización en las diversas unidades académicas (Clayson, 2009).

Finalmente, las instituciones deberían considerar que a diferencia de lo planteado en la literatura (Beran et al. 2005), este estudio señala que los docentes sí declaran utilizar los resultados de los IED para la mejora de la enseñanza; y que los directivos no están utilizando estos resultados con fines de retención o incentivos para sus equipos académicos. A diferencia de otras investigaciones (Spooren et al, 2013; Uttl, White, & González, 2017), los hallazgos de este estudio podrían indicar que los propósitos formativos del uso de los IED podrían no estar distorsionados. Sin embargo, parece que las preocupaciones de los docentes y directivos los han llevado a tener precaución con el uso de los resultados.

Conclusiones

Los resultados de este estudio permiten aportar evidencias para orientar la mejora de los IED en las instituciones de educación superior. A pesar de que la estabilidad de las puntuaciones no parece ser un problema, sí resulta relevante analizar las dimensiones y contenidos de los IED particularmente si se espera utilizar las puntuaciones de los docentes para promover la mejora de la calidad de la docencia.

En ese sentido, se podría intuir que los IED deberían centrarse más bien en reconocer una docencia que genere experiencias de aprendizaje relevantes, que una serie de dimensiones sobre lo que hace o no el docente al momento de ejercer su rol, que ha sido el enfoque que ha primado tradicionalmente.

Este análisis también debería alertar a las autoridades universitarias respecto de las diferencias disciplinares entre unidades académicas, considerando que en algunas áreas la relación de las puntuaciones IED con el rendimiento académico es más fuerte que en otras. Esto podría estar generando situaciones injustas entre el profesorado que trascienden a sus competencias profesionales.

A modo de recomendación general, las universidades e instituciones de educación superior en general deberían transitar desde la mirada de los estándares de desempeño docente, hacia un enfoque centrado en el aprendizaje. En otras palabras, reconocer aquellas experiencias de calidad que permitan el logro de los objetivos de aprendizaje previstos en la formación profesional.

Referencias

Abrami, P. & d'Apollonia, S. (1991). Multidimensional student's evaluation of teaching effectiveness-generalizability of N=1 research: comments of Marsh (1991). *Journal of Educational Psychology*, 83, 411-415. DOI: <http://dx.doi.org/10.1037/00220663.83.3.41>

American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington: DC: American Psychological Association

Beran, T., Violato, C., Kline, D., & Frideres, J. (2005). The utility of student ratings of instruction for students, faculty and administrators: a "consequential validity" study. *The Canadian Journal of Higher Education*, 2, 49-70. Recuperado de: <http://www.internationalgme.org/Resources/>

Pub s/Beran%20et%20al%20-%20Consequential%20validity.pdf

Braun, V. & Clarke V. (2008). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. DOI: <https://doi.org/10.1191/1478088706a>

Biggs. J. (2008). *Calidad del aprendizaje universitario*. Madrid, España: NARCEA.

Clayson, D. (2009). Student evaluations of teaching: Are they related to what students learn? A meta-analysis and review of the literature. *Journal of Marketing Education*, 31, 16-30. DOI: <https://doi.org/10.1177/0273475308308308>

Costin, F., Greenough, T., & Menges, R. (1971). Student Ratings of College Teaching: Reliability, Validity, and Usefulness. *Review of Educational Research*, 41(5), 511-535. DOI: <https://doi.org/10.3102/003465430410055>

Coll, C & Sánchez, E. (2008) El análisis de la interacción alumno-profesor: líneas de investigación. *Revista de Educación*, 346, 15-32. Recuperado de: http://www.revistaeducacion.mec.es/re346/re346_01.pdf

Cohen, P. (1981). Student ratings of instruction and student achievement: A metaanalysis of multisection validity studies. *Review of Educational Research*, 281-309. DOI: <https://doi.org/10.3102/0034654305103281>

García, G. (2000). Las dimensiones de la efectividad docente, validez y confiabilidad de los cuestionarios de evaluación de la docencia: síntesis de investigación internacional. En M. Rueda, & D.-B. F (Eds.), *Evaluación de la docencia. Perspectivas Actuales*. México, DF: PAIDOS.

Griffin, T., Hilton, J., P, P. & Barret, D. (2014). Correlation between point averages and student evaluation of teaching scores: taking a closer look. *Assessment & Evaluation in Higher Education*, 39, 339-348. DOI: <https://doi.org/10.1080/02602938.2013.831809>

Hogan, T. (2004). *Pruebas psicológicas: una introducción práctica*. Colombia, Bogotá: Manual Moderno.

Huybers, T. (2013). Student evaluation of teaching: the use of best-worst scaling. *Assessment & Evaluation in Higher Education*, 39(4), 496-513. DOI: <https://doi.org/10.1080/02602938.2013.851782>

Kane, M. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, 50(1), 1-73. DOI: <https://doi.org/10.1111/jedm.12000>

Marchant, J. Fauré, J. Abricot, N. (2016). Adaptación y Validación Preliminar del SPQ y el CEQ Para el Estudio de la Formación en Docencia Universitaria

en el Contexto Chile. *Psykhé*, 25(2), 1-18.
 DOI: [10.7764/psykhe.25.2.873](https://doi.org/10.7764/psykhe.25.2.873)

Moore, S. & Kuol, N. (2005) Students evaluating teachers: Exploring the importance of faculty reaction to feedback on teaching. *Teaching in Higher Education* 10(1):57-73. Doi: 10.1080/1356251052000305534

Nasser, F. & Fresko, B. (2001). Interpreting student ratings: Consultation, instructional modification, and attitudes towards course evaluation. *Studies in Educational Evaluation*, 27, 291–305. DOI: [https://doi.org/10.1016/s0191-491x\(01\)00031-1](https://doi.org/10.1016/s0191-491x(01)00031-1)

Onwuegbuzie, A., Daniel, L., & Collins, K. (2009). A meta-validation model for assessing the score- validity of student teaching evaluations. *Quality & Quantity*, 43, 197–209. DOI: <https://doi.org/10.1007/s11135-007-91124>

Onwuegbuzie, A., Witcher, A., Collins, K. Filer, J., Moore, C. (2007). Student’s Perceptions of characteristics of effective college teachers: a validity study of a teaching evaluation form using a mixed- methods analysis. *American Educational Research Journal*, 44, 113-160. DOI: <https://doi.org/10.3102/00028312062>

Ory, J., & Ryan, K. (2001). How do student ratings measure up to a new validity framework? *New Directions for Institutional Research*, 109, 27–44. DOI: <https://doi.org/10.1002/ir.2>

Padilla, J., Gómez, J., Hidalgo, D & Muñiz, J. (2007) Esquema conceptual y procedimientos para analizar la validez de las consecuencias del uso de los test. *Psicothema*, 19 (1), 173-178. Recuperado de: <http://www.psycothema.es/pdf/3345.pdf>

Penny, A. (2003). Changing the agenda for research into students’ views about university teaching: Four shortcomings of SRT research. *Teaching in Higher Education*, 8, 399-411. DOI: <https://doi.org/10.1080/13562510309396>

Ramsden P. (2003). *Learning to teach in higher education*. London: Routledge Falmer. DOI: <https://doi.org/10.4324/9780203507711>

Richardson, J. (2004). Methodological Issues in Questionnaire-Based Research on Student Learning in Higher Education. *Educational Psychology Review*, 16, 4, 347- 358. DOI: <https://doi.org/10.1007/s10648-004-0004-z>

Salazar, A. (2008). Diagnóstico preliminar sobre evaluación de la docencia universitaria. Una aproximación a la realidad en las universidades públicas y /o estatales en Chile. *Revista Iberoamericana de Evaluación Educativa*, 1(3), 68-84.

Shulman, L., (1989). Paradigmas y programas de investigación en el estudio de la enseñanza: una perspectiva contemporánea” En: M.

Wittrock (Eds). *La investigación en la enseñanza*. España, Barcelona: Paidós.

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the Validity of Student Evaluation of Teaching: The State of the Art. *Review of Educational Research*, 83, 598-642.

Webster, J. Chan W. Prosser, M. & Watkins, D. (2009). Undergraduates’ learning experience and learning process: quantitative evidence from the East. *Higher Education*, 58, 375–386. DOI: <https://doi.org/10.1007/s10734-009-9200-6>

Zabalza, M. (2009). *Competencias Docentes del Profesorado Universitario*. España, Madrid: NARCEA.

Más información en:



[@milenio_edsup1](https://www.instagram.com/milenio_edsup1)



[@milenio_EdSup](https://twitter.com/milenio_EdSup)



www.nmedsup.cl