

# Modelo de retroalimentación mediante evaluación de 360° para la docencia de pregrado en ciencias de la salud

Claudio BÁEZ-ROJAS, Karen CÓRDOVA-LEÓN, Lincoyán FERNÁNDEZ-HUERTA, Ricardo VILLAGRA-ASTUDILLO, Laura ARAVENA-CANESE

**Introducción.** La educación es un proceso dinámico que precisa evaluación constante. Los docentes necesitan un modelo de evaluación amplio que les permita observar la mayor cantidad de variables posibles, para, de esta forma, aportar a la mejora continua del proceso formativo.

**Objetivo.** Elaborar una propuesta de evaluación y retroalimentación de 360° para la docencia de educación superior en ciencias de la salud, tomando como base los componentes: evaluación por autoridades académicas, evaluación por parte de los estudiantes, autoevaluación y evaluación por pares académicos.

**Materiales y métodos.** Se aplicó la evaluación en 360° a ocho docentes de la carrera de Fonoaudiología de la Universidad San Sebastián (Chile). El instrumento se concentró en la evaluación de competencias transversales catalogadas en cuatro dimensiones: disciplinar, pedagógica, actitudinal y gestión. Se aplicaron cuatro encuestas: evaluación por jefatura, por docente par, por estudiantes y autoevaluación.

**Resultados.** La propuesta presentó alta validez de contenido, indicadores con una buena capacidad de discriminación y fiabilidad de los ítems en las distintas escalas.

**Conclusión.** Se recomienda su uso en situaciones similares donde el objetivo sea obtener una mirada integral de la calidad del servicio docente entregado.

**Palabras clave.** Enseñanza. Éxito académico. Modelos educativos. Personal educativo.

## 360-degree feedback model for undergraduate teaching in health sciences

**Introduction.** Education is a dynamic process that requires constant evaluation. Teachers need a broad evaluation model that allows them to observe as many variables as possible, thus contributing to the continuous improvement of the training process.

**Objective.** Develop a 360-degree feedback model for higher education teaching in health sciences, based on the following components: evaluation by academic authorities, evaluation by students, self-evaluation and evaluation by academic peers.

**Materials and methods.** The 360-degree feedback was applied to 8 teachers of the Speech Therapy career at the San Sebastián University (Chile). The instrument focused on the evaluation of transversal competences cataloged in four dimensions: Disciplinary, Pedagogical, Attitudinal and Management. Four surveys were applied: evaluation by headquarters, by peer teacher, by students and self-evaluation.

**Results.** The model presented high content validity, indicators with a good capacity for discrimination and reliability of the items in the different scales.

**Conclusion.** The model is recommended in similar situations where the objective is to obtain a comprehensive view of the quality of the teaching service provided.

**Key words.** Academics success. Educational models. Educational personnel. Teaching.

## Introducción

La sociedad redefine el tipo ideal de educación y va moldeando a los individuos de acuerdo con sus propios requerimientos. Sin embargo, es imperativa la necesidad de evaluar de manera sistemática y periódica toda la estructura y los componentes del

proceso de formación, con un gran esfuerzo en el recurso humano involucrado en la búsqueda de una mejor calidad [1-3]. En este proceso, los grandes cambios que suceden y sus escenarios futuros tienen que desarrollarse según las nuevas tendencias internacionales en educación, fundamentadas en la globalización y el mejoramiento de la calidad

Facultad de Ciencias de la Salud. Universidad San Sebastián (C. Báez-Rojas, L. Fernández-Huerta, R. Villagra-Astudillo, L. Aravena-Canese). Facultad de Ciencias de la Salud. Universidad de las Américas. Concepción, Chile (K. Córdova-León).

### Correspondencia:

Dra. Karen Córdova León. Facultad de Ciencias de la Salud. Universidad de las Américas. Jorge Alessandri, 1160. Concepción, Bío Bío, 4080871 Chile.

### E-mail:

kcordova@udla.cl

### Recibido:

07.01.21.

### Aceptado:

15.04.21.

### Conflicto de intereses:

No declarado.

### Competing interests:

None declared.

### Cómo citar este artículo:

Báez-Rojas C, Córdova-León K, Fernández-Huerta L, Villagra-Astudillo R, Aravena-Canese L. Modelo de retroalimentación mediante evaluación de 360° para la docencia de pregrado en ciencias de la salud. FEM 2021; 24: 173-81. doi: 10.33588/fem.244.1133.

### © 2021 FEM



Artículo open access bajo la licencia CC BY-NC-ND (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

ISSN: 2014-9832

ISSN (ed. digital): 2014-9840

y la propia cobertura, que deben acompañarse de las transformaciones pedagógicas en la enseñanza superior, a la cual también concierne la educación en ciencias de la salud [4]. Queda de manifiesto, entonces, que ya no es suficiente ser un buen profesional de la salud, sino también lograr las competencias para la profesionalización docente. De ahí que sea oportuno reflexionar acerca del papel docente en ciencias médicas y qué estrategias presenta el ejercicio responsable y adecuadamente pedagógico de su labor, tanto en el aula como en los escenarios de práctica [5], porque la sola vinculación de un profesional de la salud a un centro clínico donde se tiene un convenio asistencial-docente no es suficiente para convertirlo en docente [6-8]. Adicionalmente, los docentes de la educación superior se encuentran inmersos en la sociedad del conocimiento y de la información, cada vez más exigente [9]. Así también, el gran volumen de datos ha logrado que las estructuras de las disciplinas profesionales sean cada día más complejas, generando incluso una crisis ante las demandas de las dinámicas globalizadoras interdisciplinarias y transdisciplinarias. La futura sociedad se perfila con un alto nivel de requerimientos y desafíos, los cuales deben asumirse con responsabilidad por los docentes, quienes también deben estar preparados para los cambios que acontecen continuamente en el mundo [10]. Bajo esta premisa, es lógico pensar que para lograr una educación de calidad se requiera formar docentes de excelencia. En este sentido, una de las recomendaciones de la Unesco para fortalecer su protagonismo y corresponsabilidad con los cambios educativos es la de optimizar los sistemas de formación y evaluación docentes [10].

La evaluación del desempeño es fundamental para satisfacer diferentes necesidades organizacionales e individuales, y, si bien los procesos de reforma institucional requieren tiempo para desarrollarse, es muy oportuno detectar las experiencias significativas que permitan precisar nuevas acciones tendientes a la mejora de los procesos de enseñanza-aprendizaje mediante la evaluación del desempeño docente [11].

Una de las formas más sencillas para evaluar distintos aspectos relacionados con el desempeño docente son los cuestionarios. Pero, para brindar garantía de la utilidad y el significado de los resultados obtenidos en ellos, no basta la simple creación de una lista de preguntas y respuestas, sino que el instrumento debe estar calibrado [12]. Por tanto, el uso de cuestionarios implica necesariamente el control de la bondad de éste mediante un estudio de validez y fiabilidad. La fiabilidad de un instrumento se refie-

re a la confianza que se concede a los datos que se obtienen y está relacionada con la coherencia o la consistencia interna y la precisión de las medidas recopiladas. La validez se refiere al grado en que el instrumento mide lo que se pretende medir. Estas dos condiciones son cruciales, porque, si el instrumento es 'defectuoso', no se puede garantizar el éxito del diagnóstico, ya que el tratamiento estadístico no conseguirá transformar datos de mala calidad en buenos resultados [13].

Por lo anterior, el presente estudio tuvo como objetivo elaborar una propuesta de evaluación y retroalimentación de 360° para la docencia de educación superior en ciencias de la salud, que pretende evaluar a los docentes integralmente, al tomar como base los componentes evaluación por autoridades académicas, evaluación por parte de los estudiantes, autoevaluación y evaluación por pares académicos.

## Materiales y métodos

### Instrumento

El sistema de evaluación de 360° es una valoración por competencias que ayuda a identificar y ponderar conductas en el ejercicio profesional. Tiene su origen y aplicaciones más frecuentes en el mundo corporativo, donde se evalúan los 'comportamientos laborales' y los 'resultados' se califican anónimamente por personas de una jerarquía mayor, de una jerarquía menor y con el mismo nivel que el empleado, para proporcionar diferentes perspectivas. Este enfoque aprovecha su sabiduría colectiva para proporcionar una forma más equilibrada, completa, precisa y justa de evaluación, a diferencia del método tradicional con una única fuente. Los resultados se comparan con la autoevaluación para brindar comentarios precisos al evaluado y poder planificar mejoras específicas en su desempeño laboral para cumplir con los objetivos profesionales [14]. En algunos casos, los resultados pueden utilizarse por el supervisor para la promoción y las decisiones de índole pecuniarias, y es uno de los elementos por los cuales los evaluados pueden presentar resistencia al proceso [15].

En el ámbito de la salud, la evaluación de 360° es una herramienta que se ha utilizado ampliamente en ciencias médicas, particularmente en los últimos niveles formativos. Además de perseguir una evaluación integral del proceso, se ha usado con fines de robustecer las competencias que tributen al trabajo interdisciplinario y desarrollar habilidades de liderazgo, las cuales se ven facilitadas en gran

medida por la disponibilidad de comentarios de desempeño de alta calidad de subordinados, pares, superiores o clientes [16]. En el ejercicio docente, si bien la información de experiencias con estas estrategias es limitada, parece prometedora. De hecho, la evidencia para determinar si los profesores de ciencias de la salud realmente cambian sus conductas de enseñanza todavía es controvertida [17,18].

En esta investigación se elaboraron cuatro pautas de evaluación de la calidad docente bajo el modelo de la evaluación de 360° (una para cada actor relevante del proceso de enseñanza-aprendizaje: estudiante, docente par, jefatura y docente evaluado). Para la elaboración de estos instrumentos se realizó una búsqueda bibliográfica sobre el método de evaluación en 360° en educación y calidad docente, y, posteriormente, se procedió a seleccionar los ítems por dimensión que tenían mayor afinidad con el proyecto educativo institucional, lo que, a su vez, es coherente con las pautas internacionales de educación superior.

## Sujetos

Se aplicó la evaluación en 360° a ocho docentes de la carrera de Fonoaudiología de la sede de Concepción, responsables de las secciones teóricas de las asignaturas pertenecientes al sexto y el octavo semestres de la carrera. Se excluyó a todos los que tenían funciones directivas y que realizaban clases de forma esporádica. Se aplicaron cuatro encuestas: evaluación por jefatura (un único evaluador), evaluación por docente par (un único evaluador), evaluación por estudiantes (un curso de estudiantes) y autoevaluación; complementariamente, se aplicó una encuesta sociodemográfica. Para el caso de la evaluación por estudiantes, se promediaron las respuestas totales del curso de cada docente para obtener un indicador global. Cada encuesta está integrada por una cierta cantidad de ítems ponderados en escala de Likert, que permiten al usuario responder respecto al grado de cumplimiento de la competencia evaluada y la frecuencia con que se presenta, obteniendo así mayor objetividad en la medición del desempeño. Se eligió evaluar competencias transversales, puesto que éstas se consideran destrezas que permiten desarrollar y aplicar las capacidades propias en actividades diversas que implican factores tanto cognitivos como afectivos y comportamentales, logrando un resultado cualitativamente superior al estándar. Las preguntas se encuentran divididas en cuatro dimensiones: disciplinar, pedagógica, actitudinal y gestión. Las respuestas se recogen mediante escalas de Likert, con

una puntuación mínima de 1 y una máxima de 5, donde las puntuaciones bajas indican discrepancia y las altas corresponden a mayor acuerdo con la sentencia del ítem.

Todos los participantes firmaron un consentimiento informado donde se declaró la voluntariedad del estudio, el objetivo y los instrumentos de evaluación. También se explicitaron las condiciones y garantías de confidencialidad de la información referida por parte de los evaluados. El estudio fue aprobado por el comité de ética de la Universidad San Sebastián (Chile).

## Análisis estadístico

Se realizó un análisis de validez de contenido mediante valoración de juicio de expertos, con profesionales cualificados en el área. También se evaluó la confiabilidad mediante un análisis de consistencia interna con el objeto de dotar de significación a las preguntas del cuestionario, por lo que se calculó el coeficiente alfa de Cronbach; los valores de este coeficiente oscilan entre 0 y 1, y se consideró, como criterio general, un coeficiente aceptable cuando su valor fuera igual o superior a 0,7. Por último, se evaluó la confiabilidad mediante un análisis de la capacidad de discriminación de los ítems, de modo que se reforzara el carácter unidimensional de la prueba; para esto se calculó el índice de homogeneidad de cada ítem (el coeficiente de correlación de Pearson entre la puntuación en el ítem y la suma de las puntuaciones en los restantes ítems). Los ítems con bajos índices de homogeneidad miden algo diferente a lo que refleja la prueba en su conjunto, por lo que se pueden eliminar del cuestionario. Normalmente, se eliminan aquéllos cuyo índice de homogeneidad es menor de 0,2. No fue posible analizar la validez de constructo mediante un análisis factorial exploratorio por componentes principales debido al tamaño de la muestra.

## Resultados

### Características externas

Se evaluó a ocho docentes con edades comprendidas entre los 18 y los 38 años, con una media de  $32 \pm 3,77$ ; el 25% de la muestra correspondía al género masculino y el 75%, al femenino. Del total, el 50% eran docentes de planta (contratados) y el otro 50% eran docentes adjuntos (actividades a honorarios). La media de años de egreso de los docentes fue de  $8,12 (\pm 3,75)$  y la media de experiencia en años de

**Tabla I.** Características de la muestra.

	Mín.	Máx.	M	DE
Edad (años)	28	38	32	3,77
Experiencia de pregrado (años)	3	8	5,62	1,84
Experiencia de posgrado (años)	0	2	0,75	0,7
Tiempo de egreso de docentes (años)	4	14	8,12	3,75
			<i>n</i>	%
Género	Masculino		2	25
	Femenino		6	75
Jerarquía	No tiene		6	75
	Instructor		1	12,5
	Asistente		1	12,5
	Asociado		0	0
	Titular		0	0
Figura contractual	Planta		4	50
	Adjunto		4	50
Formación en educación	Asistencia a cursos, jornadas o seminarios		8	100
	Diplomado o posgrado		0	0
Grado de magíster	Sí		1	12,5
	No		7	87,5

DE: desviación estándar; M: media; Máx.: máximo; Mín.: mínimo.

docencia en pregrado fue de 5,62 ( $\pm 1,84$ ), y de posgrado, de 0,75 ( $\pm 0,7$ ). En cuanto a la jerarquía académica, el 75% de los docentes no se encontraba jerarquizado y el 25% restante sí lo estaba en las dos primeras categorías de segmentación docente, es decir, instructor y asistente (un 12,5% cada categoría), con un 87,5% de los docentes en posesión del grado académico de magister y un 100% con capacitaciones en docencia y pedagogía (Tabla I).

### Validez de contenido

El grupo de jueces para validar el contenido de las cuatro encuestas estuvo formado por tres expertos en el área de la educación universitaria, con men-

**Tabla II.** Estadísticas de fiabilidad de la encuesta.

	Alfa de Cronbach	Alfa de Cronbach al eliminar el ítem
Evaluación por jefatura	0,83	0,824
Evaluación por pares	0,905	0,927
Evaluación por estudiante	0,923	0,941
Autoevaluación	0,907	0,913

ciones en ciencias de la salud y evaluación curricular, todos ellos dedicados a la docencia directa de pregrado y posgrado. A todos se les envió el mismo documento, en el que se especificaba el objetivo de los instrumentos y su contenido, así como la forma en la que debían aportar su valoración sobre las dimensiones definidas, los ítems asociados a cada una de ellas y la escala de valoración de éstos. El 100% de los jueces coincidió en mantener las cuatro dimensiones, los ítems originales para las cuatro encuestas y la escala de valoración de Likert, por lo que no se consideró necesario calcular el índice de validez del contenido en su forma cuantitativa.

### Consistencia interna

El índice de confiabilidad fue aceptable para la encuesta de evaluación por jefatura (alfa de Cronbach = 0,83) y bueno para las otras escalas (alfa de Cronbach = 0,905, 0,907 y 0,923 para evaluación por pares, autoevaluación y evaluación por estudiantes, respectivamente y en orden ascendente) (Tabla II).

### Capacidad de discriminación de los ítems

El índice de fiabilidad de la encuesta completa de evaluación por jefatura que se obtiene mediante el cálculo del estadístico alfa de Cronbach fue de 0,83, que se considera aceptable (Tabla II). Este valor disminuye levemente si se eliminan los ítems 1, 4, 11, 13, 14 y 19, si se tiene en cuenta aquellos cuyo índice de homogeneidad es  $< 0,2$ ; pero el índice aumenta sólo si se prescinde del ítem 14 (Tabla III). Al eliminar este ítem y volver a realizar el análisis de fiabilidad para los datos resultantes, se obtiene un alfa de Cronbach de 0,824 (Tabla II), más baja que en el caso original, por lo que sería adecuado mantener los 20 ítems.

El índice de fiabilidad de la encuesta completa de evaluación por pares que se obtiene mediante el cál-

**Tabla III.** Valores de los índices de homogeneidad de cada ítem y del índice de fiabilidad del cuestionario al eliminar cada ítem para la encuesta de evaluación por jefatura.

	Ítem	Índice de homogeneidad	Alfa de Cronbach al eliminar el ítem
Disciplinar	Ítem 1	0	0,825
	Ítem 2	0,614	0,8
Pedagógica	Ítem 3	0,608	0,809
	Ítem 4	0,544	0,806
	Ítem 5	0,608	0,809
	Ítem 6	0,439	0,812
Actitudinal	Ítem 7	0,647	0,798
	Ítem 8	0,506	0,809
	Ítem 9	0,647	0,798
	Ítem 10	0,561	0,804
Gestión	Ítem 11	0	0,825
	Ítem 12	0,204	0,822
	Ítem 13	0	0,825
	Ítem 14	-0,197	0,833
	Ítem 15	0,204	0,822
	Ítem 16	0,204	0,822
	Ítem 17	0,328	0,822
	Ítem 18	0,317	0,818
	Ítem 19	0,098	0,824
	Ítem 20	0,69	0,794

culo del estadístico alfa de Cronbach fue alto, con un valor de 0,905 (Tabla II). Este valor disminuye si se eliminan los ítems 2, 3, 22 y 23 de forma separada, considerando un índice de homogeneidad  $< 0,2$  (Tabla IV). Al eliminar los cuatro ítems con baja homogeneidad y volver a realizar el análisis de fiabilidad para los datos resultantes, se obtiene un alfa de Cronbach de 0,927 (Tabla II), más alta que en el caso original. Después del nuevo análisis de fiabilidad, se pueden mantener los 19 ítems, dado que el valor de fiabilidad permanece.

**Tabla IV.** Valores de los índices de homogeneidad de cada ítem y del índice de fiabilidad del cuestionario al eliminar cada ítem para la encuesta de evaluación por pares.

	Ítem	Índice de homogeneidad	Alfa de Cronbach al eliminar el ítem
Disciplinar	Ítem 1	0,309	0,896
	Ítem 2	0	0,898
	Ítem 3	0	0,898
	Ítem 4	-0,221	0,908
	Ítem 5	0,938	0,876
	Ítem 6	0,922	0,892
	Ítem 7	0,755	0,889
	Ítem 8	0,492	0,893
	Ítem 9	0,946	0,884
	Ítem 10	0,321	0,898
Pedagógica	Ítem 11	0,28	0,896
	Ítem 12	0,791	0,888
	Ítem 13	0,639	0,888
	Ítem 14	0,506	0,893
	Ítem 15	0,563	0,891
	Ítem 16	0,776	0,885
	Ítem 17	0,636	0,888
Actitudinal	Ítem 18	0,755	0,889
	Ítem 19	0,791	0,888
	Ítem 20	0,662	0,892
	Ítem 21	0,882	0,88
Gestión	Ítem 22	0	0,898
	Ítem 23	0	0,898

El índice de fiabilidad de la encuesta completa de evaluación por estudiantes que se obtiene mediante el cálculo del estadístico alfa de Cronbach fue alto, con un valor de 0,923 (Tabla II). Este valor aumenta levemente si se eliminan los ítems 2, 7, 8, 9, 12 y 21, si se tiene en cuenta el índice de homogeneidad  $<$

**Tabla V.** Valores de los índices de homogeneidad de cada ítem y del índice de fiabilidad del cuestionario al eliminar cada ítem para la encuesta de evaluación por estudiantes.

	Ítem	Índice de homogeneidad	Alfa de Cronbach al eliminar el ítem
Disciplinar	Ítem 1	0,597	0,922
	Ítem 2	0,14	0,924
	Ítem 3	0,289	0,923
	Ítem 4	0,756	0,917
	Ítem 5	0,554	0,92
	Ítem 6	0,738	0,92
	Ítem 7	0,269	0,924
	Ítem 8	-0,161	0,926
	Ítem 9	0,26	0,925
	Ítem 10	0,305	0,923
	Ítem 11	0,718	0,918
Pedagógica	Ítem 12	-0,199	0,925
	Ítem 13	0,607	0,919
	Ítem 14	0,306	0,923
	Ítem 15	0,962	0,916
	Ítem 16	0,722	0,917
	Ítem 17	0,837	0,915
	Ítem 18	0,951	0,912
	Ítem 19	0,924	0,913
	Ítem 20	0,409	0,922
	Ítem 21	0,617	0,924
Actitudinal	Ítem 22	0,834	0,918
	Ítem 23	0,926	0,919
	Ítem 24	0,723	0,921
	Ítem 25	0,432	0,922
	Ítem 26	0,817	0,92
	Ítem 27	0,469	0,922
Gestión	Ítem 28	0,565	0,922
	Ítem 29	0,828	0,915
	Ítem 30	0,491	0,921
	Ítem 31	0,386	0,922

0,2 (Tabla V). Al eliminar los seis ítems indicados y volver a realizar el análisis de fiabilidad para los datos resultantes, se obtiene un alfa de Cronbach de 0,941, más alta que en el caso original (Tabla II). Después del nuevo análisis de fiabilidad, se pueden mantener los 25 ítems, dado que el valor de fiabilidad permanece.

El índice de fiabilidad de la encuesta completa de autoevaluación que se obtiene mediante el cálculo del estadístico alfa de Cronbach fue alto, con un valor de 0,907 (Tabla II). Este valor aumenta levemente si se elimina el ítem 32, y los ítems 2, 4, 11 y 34 si se tiene en cuenta el índice de homogeneidad  $< 0,2$ ; por otro lado, si se eliminan los ítems 9, 30 y 31, según el índice de homogeneidad, la fiabilidad permanece (Tabla V). Al eliminar los ocho ítems indicados y volver a realizar el análisis de fiabilidad para los datos resultantes, se obtiene un alfa de Cronbach de 0,913, más alta que en el caso original (Tabla II). Después del nuevo análisis de fiabilidad, se pueden mantener los 40 ítems, dado que el valor de fiabilidad permanece.

## Discusión

La encuesta de evaluación por jefatura ofrece un buen índice de fiabilidad cuando se elimina una de las preguntas relacionadas con el dominio disciplinar y cuatro preguntas del dominio de gestión. El ítem 14 aumenta en mayor proporción la fiabilidad; éste alude a la entrega oportuna de las evaluaciones a los estudiantes, por lo que se podría prescindir de ellas a la hora de medir las dificultades del docente y su gestión de tiempo para la entrega de evaluaciones. Probablemente, la jefatura no sea el informante más idóneo para evaluar este ítem, pero sí podrían serlo el estudiante y el propio docente, ya que el análisis estadístico no sugirió prescindir de las preguntas similares en las otras escalas. Aunque se debe tener en cuenta que la evidencia ya ha denotado discrepancia entre estas últimas figuras, ya que el estudiante valora mayormente la puntualidad de la entrega de una retroalimentación y el docente prioriza los comentarios informativos [19,20]. Sobre el tiempo en específico, se ha descrito que los estudiantes no tienen conocimiento de los planes de trabajo o los compromisos de sus docentes, por lo que tienen una comprensión limitada de las demandas competitivas en el tiempo de éstos, más allá de la enseñanza, y podrían ser más exigentes en este sentido [21]. En este último punto, es de suponer que la jefatura podría tener una mayor comprensión, sobre todo en los casos de docentes de

planta, donde gran parte del tiempo está dedicado a actividades de gestión y no académicas.

Para la encuesta de evaluación por pares, al eliminar alguno de los ítems de baja homogeneidad se esperaba que disminuyera el índice global de fiabilidad, pero ocurre lo contrario al eliminar estas cuatro preguntas. Dos de estos ítems corresponden al dominio disciplinar y dos de ellos al dominio de gestión (los únicos pertenecientes al dominio), donde la pertinencia de estos ítems podría considerarse al contrastar la evidencia, que enuncia que la evaluación por pares crea oportunidades para mejorar la calidad de la enseñanza y la satisfacción de los docentes, y para promover una cultura de excelencia en la enseñanza. Sin embargo, programas de revisión por pares requieren objetivos claros, infraestructura para la coordinación del programa y, sobre todo, capacitación de los revisores pares. De hecho, se recomienda un grupo central de revisores por pares para promover una observación más objetiva [22]. Bajo esta premisa, es lógico pensar que un único evaluador par podría no ser suficiente para brindar una respuesta de alta fiabilidad para el dominio de gestión. La bibliografía también revela que la resistencia del cuerpo docente a la evaluación por pares es frecuente y puede dificultar su uso. La falta de apelación se atribuye a las percepciones de los docentes, al considerar que este tipo de evaluaciones pueden estar sesgadas, y que un revisor o una observación no proporcionan datos suficientes y que no son apropiados para la evaluación sumativa [23]. Para conciliar estas preocupaciones, la observación múltiple entre pares es imprescindible. Pero, además, la bibliografía sugiere que los docentes sean incluidos en la planificación del proceso y la creación de las herramientas para la medición, reconozcan su contribución al mejoramiento de sus colegas y visualicen este insumo como mejora continua a su propio desarrollo profesional [24].

La evaluación por estudiantes es la herramienta más común para evaluar el desempeño docente, y se considera el método más eficaz y confiable, aunque controvertido [25,26]. La encuesta propuesta para ser desarrollada por estudiantes presentó un alto índice de fiabilidad, pero se sugiere considerar eliminar algunos ítems, de los cuales la mayoría se relacionan directamente con el dominio disciplinar. Probablemente, estos resultados tengan que ver con la propia competencia de los estudiantes y su heterogeneidad por la cantidad de fuentes para evaluar papeles de enseñanza como los que involucran el diseño del curso (objetivos, contenido, métodos y evaluación) o la práctica de calificación en la evaluación. Aunque, por otro lado, en términos

**Tabla VI.** Valores de los índices de homogeneidad de cada ítem y del índice de fiabilidad del cuestionario al eliminar cada ítem para la encuesta de autoevaluación.

	Ítem	Índice de homogeneidad	Alfa de Cronbach al eliminar el ítem
Disciplinar	Ítem 1	0,568	0,902
	Ítem 2	0,096	0,908
	Ítem 3	0,336	0,905
	Ítem 4	-0,147	0,909
	Ítem 5	0,463	0,903
	Ítem 6	-0,106	0,909
	Ítem 7	0,882	0,899
	Ítem 8	0,508	0,903
	Ítem 9	0	0,907
	Ítem 10	0,518	0,903
	Ítem 11	-0,014	0,908
	Ítem 12	0,336	0,905
	Ítem 13	0,785	0,901
	Ítem 14	0,739	0,900
	Ítem 15	0,785	0,900
Pedagógica	Ítem 16	0,893	0,898
	Ítem 17	0,459	0,905
	Ítem 18	0,601	0,902
	Ítem 19	0,779	0,900
	Ítem 20	0,448	0,904
	Ítem 21	0,795	0,898
	Ítem 22	0,256	0,906
	Ítem 23	0,251	0,906
	Ítem 24	0,548	0,902
	Ítem 25	0,632	0,902
	Ítem 26	0,357	0,905
	Ítem 27	0,578	0,902
	Ítem 28	0,317	0,905

**Tabla VI.** Valores de los índices de homogeneidad de cada ítem y del índice de fiabilidad del cuestionario al eliminar cada ítem para la encuesta de autoevaluación (*cont.*).

	Ítem	Índice de homogeneidad	Alfa de Cronbach al eliminar el ítem
Actitudinal	Ítem 29	0,591	0,903
	Ítem 30	0,016	0,907
	Ítem 31	0,016	0,907
	Ítem 32	-0,436	0,909
	Ítem 33	0,507	0,904
	Ítem 34	0,521	0,903
	Ítem 35	0,615	0,901
	Ítem 36	0,566	0,903
	Ítem 37	0,116	0,907
	Ítem 38	-0,112	0,910
Gestión	Ítem 39	0,839	0,901
	Ítem 40	0,839	0,901
	Ítem 41	0,021	0,911
	Ítem 42	0,161	0,906
	Ítem 43	0,222	0,906
	Ítem 44	0,108	0,908
	Ítem 45	0,491	0,905
	Ítem 46	0,661	0,902

de la calidad en la entrega de la instrucción, generalmente se acepta que sólo los estudiantes están en condiciones de proporcionar una buena retroalimentación [26]. Se ha demostrado que ciertas variables propias de los docentes (género, edad, experiencia docente, personalidad y productividad en investigación), las variables propias de los estudiantes (género, edad, nivel, promedio de calificaciones y personalidad), las variables propias del curso (tamaño de la clase y hora del día en que se imparte la clase) y las variables administrativas (tiempo del módulo durante el período) generalmente no afectan a las evaluaciones realizadas por los estudiantes sobre la calidad de la enseñanza [27].

Para el instrumento de autoevaluación, se encontró una alta fiabilidad; sin embargo, fue mayor la falta de homogeneidad por ítems. Del análisis de fiabilidad, se desprende que es necesario revisar la pertinencia de ocho ítems en total, con énfasis en el dominio actitudinal, y esto podría deberse, en parte, a la estructura gramatical o al sentido de la afirmación. De hecho, es más probable encontrar disparidad de respuestas en instrumentos de autoevaluación que en otro tipo [28]. A pesar de esto, Danielson (1996) recomendó la autoevaluación como la herramienta más poderosa para medir la calidad docente y, aunque existe la posibilidad de que los maestros experimentados se califiquen más alto en los indicadores de efectividad de la enseñanza [29], la evidencia del proceso de autoevaluación puede proporcionar apoyo para el docente sobre sus actividades en el aula y ser una buena herramienta para visualizar su desempeño, que es muy difícil de conseguir de cualquier otra fuente [20,30].

Evaluar las contribuciones de los docentes al aprendizaje de los estudiantes no es una tarea fácil en ningún entorno, y se vuelve aún más difícil en materias como la educación en ciencias de la salud cuando llegan a niveles de profesionalización de la carrera, más aún cuando se imparten en un entorno diferente al de un aula tradicional. El objetivo principal de la evaluación docente debe ser proporcionarles herramientas para ser más efectivos, y, en este sentido, el sistema de evaluación debe facilitar la retroalimentación objetiva y específica del contenido a los docentes, con el fin de informar de su práctica y guiarlos para mejorar su efectividad; en este sentido, una evaluación de 360° cobra mayor relevancia. A saber, si los agentes clave del proceso están genuinamente interesados en mejorar la calidad de su enseñanza, entonces la responsabilidad recae en ellos, y es importante que todos sean partícipes de los procesos evaluativos [31].

Una limitación del estudio es que la evidencia actual no es suficiente para afirmar que los profesores de ciencias de la salud realmente cambiarán sus prácticas de enseñanza utilizando el método de evaluación de 360° [32]. En este sentido, los resultados controvertidos también limitaron la comparación con los propios resultados del estudio. Investigaciones futuras podrían centrarse en qué nivel ocurren los cambios o ajustes y cuál es la durabilidad de estos cambios en el quehacer docente.

## Conclusiones

La actual propuesta de un modelo de evaluación



de 360° presenta validez de contenido, indicadores de aceptable a buena fiabilidad en las distintas escalas y una buena capacidad de discriminación de los ítems de cada una de ellas. Se recomienda su uso en situaciones similares donde el objetivo sea obtener una mirada integral de la calidad del servicio docente entregado. Dicho lo anterior, se sugiere evaluar la pertinencia de los ítems que se propuso prescindir, y siempre acompañar el análisis descriptivo con uno de validez y confiabilidad en las muestras sobre las que se apliquen los instrumentos.

### Bibliografía

1. Retrepo Gómez B. Tendencias actuales en la educación superior: rumbos del mundo y rumbos del país. *Revista Educación y Pedagogía* 2006; 18: 79-90.
2. Hernández de Menéndez M, Morales-Menéndez R. Current trends in competency based education. *World Journal of Engineering and Technology* 2016; 04: 193-9.
3. Altbach PG, Reisberg L, Rumbley LE. Tracking a global academic revolution. *Change: The Magazine of Higher Learning* 2010; 42: 30-9.
4. Centeno A. Las ciencias del aprendizaje: su potencial influencia en la enseñanza de la medicina. *Rev Argent Educ Médica* 2014; 6: 43-4.
5. Manso Martínez JM. Reflexiones sobre un nuevo modelo de profesor universitario en Medicina. Una visión crítica de la Enseñanza Médica; 2001. Segundo Congreso Virtual de Cardiología. Buenos Aires: Federación Argentina de Cardiología; 2001.
6. Sutkin G, Wagner E, Harris I, Schiffer R. What Makes a good clinical teacher in medicine? A review of the literature. *Academic Medicine* 2008; 83: 452-66.
7. Fluit CRMG, Bolhuis S, Grol R, Laan R, Wensing M. Assessing the quality of clinical teachers. *Journal of General Internal Medicine* 2010; 25: 1337-45.
8. Bearman ML, Tai JH-M, Kent FM, Edouard V, Nestel DF, Molloy E. Identifying the learning priorities of clinical supervisors. Abstract from International Clinical Skills Conference. 2017. Tuscany, Italy.
9. Ching G. A literature review on the student evaluation of teaching. *Higher Education Evaluation and Development* 2018; 12: 63-84.
10. Unesco. Secretariat. World education report 1998: teachers and teaching in a changing world. Unesco, Paris, France; 1998.
11. Sifuentes Ocegueda ATS, Castellanos CAH, García PDJC. Propuesta de un modelo de evaluación del desempeño de 360° para el personal docente de Educación Superior. *UNIANDES* 2016; 3: 197-211.
12. Martínez JF, Schweig J, Goldschmidt P. Approaches for combining multiple measures of teacher performance. educational evaluation and policy analysis. *American Educational Research Association* 2016; 38: 738-56.
13. Lacave Rodero C, Molina Diaz AI, Fernández Guerrero M, Redondo Duque MA. Análisis de la fiabilidad y validez de un cuestionario docente. *ReVisión* 2016; 9: 1-14.
14. Tee D, Ahmed P. 360 degree feedback: an integrative framework for learning and assessment. *Teaching in Higher Education* 2014; 19: 579-91.
15. Cousar M, Huang M, Sebro R, Levin D, Prabhakar H. Too scared to teach? The unintended impact of 360-degree feedback on resident education. *Current problems in diagnostic radiology* 2020; 49: 239-42.
16. Garman A, Tyler J, Darnall J, Lerner W. Development and validation of a 360-degree-feedback instrument for healthcare administrators. *Journal of Healthcare Management* 2004; 49: 307.
17. Fluit C, Bolhuis S, Klaassen T, de Visser M, Grol R, Laan R, et al. Residents provide feedback to their clinical teachers: reflection through dialogue. *Medical Teacher* 2013; 35: e1485-92.
18. Fluit C, Feskens R, Bolhuis S, Grol R, Wensing M, Laan R. Repeated evaluations of the quality of clinical teaching by residents. *Perspectives on Medical Education* 2013; 2: 87-94.
19. Weaver MR. Do students value feedback? Student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education* 2006; 31: 379-94.
20. Mulliner E, Tucker M. Feedback on feedback practice: perceptions of students and academics. *Assessment & Evaluation in Higher Education* 2015; 42: 266-88.
21. Mahmood F, Hope D, Cameron H. Medical students' understanding of cost effectiveness in feedback delivery. *MedEdPublish* 2019; 8: 11-2.
22. Hyland KM, Dhaliwal G, Goldberg AN, Chen L, Land K, Wamsley M. Peer review of teaching: insights from a 10-year experience. *Medical Science Educator* 2018; 28: 675-81.
23. Berk RA, Naumann PL, Appling SE. Beyond student ratings: peer observation of classroom and clinical teaching. *Int J Nurs Educ Scholarsh* 2004; 1: article 10.
24. Blauvelt MJ, Erickson CL, Davenport NC, Spath ML. Say yes to peer review. *nurse educator*. Ovid Technologies 2012; 37: 126-30.
25. Hammonds F, Mariano GJ, Ammons G, Chambers S. Student evaluations of teaching: improving teaching quality in higher education. *Perspectives: Policy and Practice in Higher Education* 2016; 21: 26-33.
26. Husain M, Khan S. Students' feedback: an effective tool in teachers' evaluation system. *Int J Appl Basic Med Res* 2016; 6: 178-81.
27. Cohen PA, McKeachie WJ. The role of colleagues in the evaluation of college teaching. *Improving College and University Teaching* 1980; 28: 147-54.
28. Cashin WE. Student ratings of teaching: a summary of the research. *IDEA Paper No 20*; 1988.
29. Danielson C. The handbook for enhancing professional practice: using the framework for teaching in your school. Alexandria, VA, EE.UU.; ASCD; 2008.
30. Akram M, Zepeda SJ. Development and validation of a teacher self-assessment instrument. *Journal of Research & Reflections in Education* 2015; 9: 134-48.
31. Hunt K, Gurvitch R, Lund JL. Teacher evaluation: done to you or with you? *Journal of Physical Education, Recreation & Dance* 2016; 87: 21-7.
32. Boerboom T, Stalmeijer R, Dolmans D, Jaarsma D. How feedback can foster professional growth of teachers in the clinical workplace: a review of the literature. *Studies in Educational Evaluation* 2015; 46: 47-52.