

# Diseño y validación de un instrumento para la evaluación del desempeño docente en la educación superior

Gustavo Rivas-Martínez, Luis Ortiz y Fátima Bañuelos-Gómez

## RESUMEN

Se presenta un instrumento validado para la evaluación del desempeño docente en la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de Asunción, Paraguay. La validez de contenido de los ítems fue establecida mediante juicio de expertos, empleando el índice V de Aiken para valorar la pertinencia, coherencia y claridad de cada uno. La estructura teórica del cuestionario fue confirmada a través de un Análisis Factorial Confirmatorio, lo que respalda su consistencia interna y solidez conceptual. El instrumento final quedó conformado por 21 ítems distribuidos en cuatro dimensiones clave del quehacer docente: Responsabilidad, Didáctica, Evaluación y Promoción del Aprendizaje. Además, se presenta un ejemplo de aplicación real del instrumento validado, correspondiente a la evaluación docente del segundo semestre de 2024, que evidencia su utilidad para generar informes individuales y colectivos orientados a la mejora continua.

**Palabras clave:** percepción, evaluación de profesores, instrumentos de evaluación, validez, Paraguay.

### Gustavo-Ignacio Rivas-Martínez

[grivas@facen.una.py](mailto:grivas@facen.una.py)

Paraguay. Doctor en Estadística e Investigación Operativa por la Universidad de Sevilla, España. Docente investigador de tiempo completo en la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de Asunción, Paraguay. Líneas de investigación: métodos cuantitativos en investigación educativa; educación inclusiva, y adecuaciones curriculares y sus determinantes. ORCID: <https://orcid.org/0000-0002-5751-6310>.

### Luis Ortiz

[luis\\_ortiz@facso.una.py](mailto:luis_ortiz@facso.una.py)

Paraguay. Doctor en Sociología por la École des Hautes Études en Sciences Sociales, de París, Francia. Docente investigador de la Facultad de Ciencias Sociales (FACSO)-Universidad Nacional de Asunción, Paraguay. Líneas de investigación: estratificación y desigualdad social, sistema educativo y políticas públicas. ORCID: <https://orcid.org/0000-0002-1902-8834>.

### Fátima Bañuelos-Gómez

[fbanuelos@founa.edu.py](mailto:fbanuelos@founa.edu.py)

Paraguay. Magister en Metodología de la Investigación por la Universidad Iberoamericana del Paraguay, Paraguay. Profesional Estadístico e investigadora en la Universidad Nacional de Asunción, Paraguay. Líneas de investigación: estilos de aprendizaje y rendimiento académico; gestión del conocimiento; tecnologías de información y comunicación, y habilidades investigativas. ORCID: <https://orcid.org/0000-0003-3670-0941>.



## **Design e validação de um instrumento para avaliação do desempenho docente no ensino superior**

### **RESUMO**

Foi apresentado um instrumento validado para avaliação do desempenho docente na Faculdade de Ciências Exatas e Naturais da Universidade Nacional de Assunção, Paraguai. A validação do conteúdo dos itens foi confirmada pelo julgamento de especialistas, utilizando o cálculo do V de Aiken para avaliar a relevância, coerção e clareza. Além disso, o modelo de avaliação foi validado por meio da Análise Fatorial Confirmatória. O instrumento é composto por 21 itens, distribuídos em quatro dimensões: Responsabilidade, Didática, Avaliação e Promoção da Aprendizagem. Além disso, apresenta um exemplo de aplicação real do instrumento validado, correspondente à avaliação do ensino do segundo semestre de 2024, que evidencia sua utilidade para gerar relatórios individuais e coletivamente voltados à melhoria contínua.

**Palavras chave:** percepção, avaliação de professores, instrumentos de avaliação, validade, Paraguai.

---

## **Development and validation of an instrument for evaluating teaching performance in higher education**

### **ABSTRACT**

This paper introduces a validated instrument for teaching performance assessment at the Faculty of Exact and Natural Sciences at the National University of Asunción, Paraguay. The content validity of the items was confirmed through expert judgment, using Aiken's V calculation to assess relevance, coherence, and clarity. Additionally, the evaluation model was validated through Confirmatory Factor Analysis. The instrument comprises 21 items distributed across four dimensions: Responsibility, Didactics, Evaluation, and Promotion of Learning. In addition, a real-world application of the validated instrument is presented, corresponding to the teaching evaluation of the second semester of 2024, which demonstrates its usefulness for generating individual and collective reports oriented towards continuous improvement.

**Keywords:** perception, teacher evaluation, assessment instruments, validity, Paraguay.

**Recepción:** 05/09/24. **Aprobación:** 08/10/25.

## Introducción

El ejercicio docente involucra diversas actividades y responsabilidades en el proceso de enseñanza-aprendizaje orientadas al cumplimiento de los objetivos educativos institucionales y sociales. En este sentido, la evaluación del desempeño docente consiste en valorar el grado de cumplimiento de los procesos encaminados al logro de dichos objetivos, siendo crucial la retroalimentación para la mejora continua de las prácticas pedagógicas movilizadas (De Chaparro *et al.*, 2008). Además, estudios recientes muestran una asociación entre el desempeño docente y los logros de aprendizaje, lo que enfatiza la necesidad de fortalecer continuamente las competencias pedagógicas para apuntalar la calidad educativa (Gonzales, 2022).

En América Latina, la calidad educativa está estrechamente asociada con el desempeño docente, factor humano central en la eficacia del proceso de enseñanza-aprendizaje. A pesar de los avances significativos en la región, persisten desafíos ligados a la calidad y la equidad, así como a la formación docente, que presenta limitaciones, las cuales requieren atención frente a los vertiginosos cambios sociales y tecnológicos del siglo XXI (Escribano Hervis, 2018).

Es importante insistir que la evaluación del desempeño docente no se limita a la medición de los resultados académicos, sino que implica también aspectos pedagógicos, metodológicos, comunicativos e institucionales, los cuales repercuten en la experiencia de aprendizaje del estudiante (Ley-Leyva y Espinoza-Freire, 2021). No obstante, estos aspectos dependen de la validez y confiabilidad de los instrumentos utilizados para la medición. La rigurosidad en el diseño y validación de esas herramientas es crucial para asegurar que los resultados obtenidos sean consistentes y pertinentes en la toma de decisiones.

El punto de vista del estudiante es una parte importante en la evaluación del desempeño, ya que puede evidenciar algunas conductas del profesorado relacionadas con su estilo de aprendizaje lo que

permite un mayor interés por lo que se enseña (Valentín-Martínez y Mayor-Ruiz, 2023). Por ello, el uso de instrumentos de evaluación de desempeño de los docentes desde la perspectiva del estudiante permite una mayor comprensión de la enseñanza en el aula (Henríquez Ritchie *et al.*, 2024).

Aunque las instituciones de educación superior cuentan con herramientas para evaluar el desempeño de su plantel docente, a menudo los instrumentos no son expresión de un abordaje científico riguroso. En América Latina existe aún una notable carencia de estudios que aborden el diseño, la calibración y la validación de dichos instrumentos, lo que se traduce en problemas de rigurosidad de las evaluaciones mismas, como de los estudios sobre la evaluación del desempeño docente. Esto es particularmente preocupante, por lo que Manzi *et al.* (2021) señalan que muchos de los esfuerzos de evaluación no involucran una validación adecuada de los instrumentos utilizados, lo cual compromete la fiabilidad y utilidad de la información.

Si bien, más adelante, se proporcionarán detalles sobre el contexto de la investigación, es necesario adelantar que este estudio se enfoca en el desarrollo y validación de un instrumento específico de evaluación del desempeño docente, que puede considerarse un caso particular de tratamiento de dimensiones y variables, así como de contraste de formulaciones de diseño, con el fin de extrapolar sus resultados a otras esferas educativas. Se trata de la evaluación del desempeño docente en la Facultad de Ciencias Exactas y Naturales (FACEN) de la Universidad Nacional de Asunción (UNA), Paraguay, cuyo objetivo principal es el desarrollo de una herramienta válida y confiable para la evaluación efectiva del desempeño docente en esta institución educativa.

Un instrumento de medida, entendido como una técnica para cuantificar indirectamente un constructo (Hernández y Pascual, 2018), debe desarrollarse y validarse según los objetivos de la investigación y las características de la población. Su validez no es



absoluta, sino que depende de estos factores y se evalúa mediante métodos y criterios específicos (Soriano Rodríguez, 2014).

### **Marco situacional de la investigación**

En Paraguay, la creación de la Agencia Nacional de Evaluación y Acreditación de la Educación Superior (ANEAES) en 2003 y la promulgación de la Ley núm. 4995 en 2013 marcaron un hito en la educación superior. La ANEAES evalúa y acredita la calidad académica de las instituciones y programas, además, elabora diagnósticos que fortalecen los procesos educativos.

Varias unidades académicas, entre ellas la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de Asunción, se encuentran actualmente en procesos de acreditación, 2024. Una estrategia clave para mejorar la calidad educativa es sistematizar los resultados de la Encuesta de Percepción Estudiantil, sobre el Desempeño Docente. Sin embargo, los instrumentos utilizados deben validarse rigurosamente para garantizar resultados fiables y planes de mejora basados en evidencia.

La FACEN implementó, en 2014, un instrumento de evaluación docente basado en la percepción estudiantil desarrollado, internamente, por áreas administrativas. No obstante, su validación científica continúa siendo una necesidad. Para obtener conclusiones válidas, los instrumentos deben ser legítimos y fiables, midiendo con precisión lo que se pretende evaluar y ofreciendo resultados consistentes. Ante la falta de un instrumento adecuado, surge la necesidad de diseñar y validar uno nuevo.

El presente estudio desarrolla, valida y aplica un instrumento para evaluar la percepción estudiantil del desempeño docente en la educación superior. Éste se estructura en cuatro dimensiones fundamentales: Responsabilidad, Didáctica, Evaluación y Promoción del Aprendizaje, respaldadas por modelos teóricos que destacan su relevancia en la calidad educativa (Pacheco Cámara *et al.*, 2018; Stronge,

2018; Uljens, 2004; Mohammadi, 2021; Divoky y Rothermel, 1988).

El objetivo principal de este estudio es diseñar, validar y aplicar un instrumento que capture estas dimensiones en la evaluación del desempeño docente en la FACEN-UNA. Al validar y aplicar este instrumento se busca ofrecer una herramienta adaptable a otras universidades en Iberoamérica, contribuyendo así al fortalecimiento de los sistemas de evaluación educativa.

### **Marco teórico**

#### **Validez de contenido**

La validez de contenido se refiere al grado en que una prueba representa adecuadamente el dominio o constructo que pretende medir. En indicadores objetivos, como las habilidades matemáticas, su evaluación resulta sencilla; sin embargo, cuando se trata de variables psicológicas o categorías difíciles de implementar, se recurre al juicio de expertos para determinar si los ítems son adecuados (Escobar-Pérez y Cuervo-Martínez, 2008).

La evaluación del desempeño docente no puede separarse del contexto institucional, pues los distintos niveles educativos, asignaturas y entornos requieren considerar factores específicos que influyen en la validez de la evaluación. Este enfoque asegura que los procesos evaluativos se adapten a las particularidades de cada contexto.

La validez garantiza el rigor científico y la pertinencia de las decisiones, adaptándose a las características y necesidades de cada situación. Su naturaleza es multidimensional, ya que los procedimientos de validación varían según el objeto y el contexto de evaluación. Desde perspectivas no positivistas, se entiende también como un concepto dependiente del marco teórico y metodológico de la investigación (Jornet Melià *et al.*, 2020).

En la evaluación docente, la validez puede entenderse como el grado de representatividad y utilidad de los resultados, siempre que se hayan considerado

los elementos del perfil docente y las condiciones reales del proceso educativo.

Para su medición, uno de los métodos más utilizados es el coeficiente V de Aiken (1980), que cuantifica el nivel de acuerdo entre expertos respecto a la relevancia de los ítems. Su valor oscila entre 0 y 1, donde 1 indica acuerdo total. La interpretación se basa en su magnitud y significancia estadística, con valores críticos establecidos por Aiken (1985). La ecuación fue posteriormente modificada algebraicamente por Penfield y Giacobbi (2004).

$$V = \frac{\bar{x} - l}{k} \quad (1)$$

donde  $\bar{x}$  representa la puntuación media de los jueces por el ítem considerado,  $l$  es la puntuación mínima posible, y  $k$  es la puntuación máxima posible.

El intervalo de confianza (IC)  $1-\alpha$ ,  $0 < \alpha < 1$  para  $V$  viene dado por:

$$L_{inf} = \frac{2nkV + z_{(1-\alpha/2)}^2 - z_{(1-\alpha/2)}\sqrt{4nkV(1-V) + z^2}}{2(nk + z_{(1-\alpha/2)}^2)} \quad (2)$$

$$L_{sup} = \frac{2nkV + z_{(1-\alpha/2)}^2 + z_{(1-\alpha/2)}\sqrt{4nkV(1-V) + z^2}}{2(nk + z_{(1-\alpha/2)}^2)} \quad (3)$$

Donde  $L_{inf}$  y  $L_{sup}$  son los límites inferior y superior del IC para  $V$ ,  $z_{(1-\alpha/2)}$  corresponde al percentil  $1-\alpha/2$  de la normal estándar,  $n$  es la cantidad total de jueces/expertos.

### Consistencia interna

La consistencia en la evaluación implica categorías como pertinencia, coherencia, cobertura y reiteración, además de aspectos cuantitativos vinculados a la implementación curricular. El análisis de consistencia interna busca detectar desviaciones de diseño o aplicación que obstaculicen el proceso formativo, promoviendo un ciclo de análisis y mejora del instrumento.

La confiabilidad se entiende como la estabilidad de las puntuaciones obtenidas por los mismos individuos en diferentes aplicaciones o condiciones equivalentes (Anastasi y Urbina, 1998). Indica en qué medida las diferencias observadas reflejan verdaderas variaciones en la característica evaluada y no errores aleatorios. Su estimación permite calcular la proporción de la varianza total atribuible al error (Cappello *et al.*, 2004). En este sentido, las distintas formas de confiabilidad pueden expresarse mediante coeficientes de correlación que reflejan la consistencia entre puntuaciones independientes.

El coeficiente de fiabilidad representa el grado de confianza en las puntuaciones observadas como estimaciones de las verdaderas puntuaciones, midiendo el impacto del error aleatorio (Carmines y Zeller, 1979). La consistencia interna se evalúa generalmente mediante el coeficiente alfa de Cronbach, basado en la covariación entre los ítems del instrumento. Este indicador mide hasta qué punto los ítems evalúan un mismo constructo. El alfa de Cronbach es ampliamente utilizado en la validación de instrumentos en ciencias sociales, educación y salud (McNeish y Wolf, 2020). Un valor igual o superior a 0.70 suele considerarse aceptable, aunque puede variar según el contexto y los objetivos del estudio. Feldt (1965) propuso el uso de intervalos de confianza para este coeficiente, proporcionando una estimación más precisa de la fiabilidad del instrumento.

### Validez de constructo

La validez del método se refiere a la precisión con que se miden las variables de una prueba, siendo válida sólo si evalúa aquello para lo que fue construida (Magnusson, 1966). Representa el grado de correspondencia entre las puntuaciones obtenidas y las de la variable criterio, con un nivel conocido de certeza, especialmente cuando la relación entre ambas no es evidente y requiere verificación (Bartolomé, 1983). Su propósito es contrastar hipótesis sobre las relaciones entre variables derivadas de la teoría, fortaleciendo la



confianza en el instrumento y validando la red conceptual asociada al constructo. Según Cronbach y Meehl (1955), un constructo sólo puede considerarse medible si se define dentro de un marco teórico que relacione sus conceptos con observables. Este proceso suele ser inductivo, ya que la estructura conceptual puede redefinirse a partir de los datos.

La validez de constructo garantiza que un instrumento mida efectivamente el constructo propuesto, asegurando la pertinencia y precisión de las inferencias (Kline, 2015). En investigación educativa, esta validez se evalúa comparando la estructura teórica del instrumento con la estructura empírica obtenida mediante análisis factorial (Reynolds *et al.*, 2021). El análisis factorial exploratorio identifica patrones subyacentes, mientras que el confirmatorio (AFC) contrasta un modelo teórico previamente definido (Brown, 2015).

El AFC, técnica avanzada de validación de constructos, permite comprobar si los datos empíricos se ajustan a la estructura teórica propuesta (Schumacker y Lomax, 2004). En la evaluación del desempeño docente, verifica si los ítems se agrupan coherentemente según las dimensiones teóricas, confirmando que el instrumento mide los constructos esperados (Hoyle, 2012). La calidad del ajuste se determina mediante índices como CFI, TLI y RMSEA, cuyos valores aceptables respaldan la adecuación del modelo teórico a los datos (Byrne, 2012).

### **Evaluación del desempeño docente**

La evaluación del desempeño se fundamenta en modelos orientados a objetivos, donde los resultados se comparan con las metas para determinar su grado de logro. Sin embargo, estos modelos presentan limitaciones, ya que la información obtenida suele ser extemporánea respecto a los procesos formativos en curso (Mora, 2004). Frente a ello, surgen modelos centrados en el proceso, como el Contexto, Insumo, Proceso y Producto (CIPP) de Stufflebeam (2003) y el modelo de Evaluación de Procesos de Cronbach,

que amplían la mirada más allá de los resultados.

Otro enfoque considera la transacción y la interpretación, incorporando la perspectiva de los sujetos y el carácter cualitativo de la evaluación. En este caso, los criterios se establecen de manera intersubjetiva, buscando identificar factores que favorezcan los resultados esperados (Stufflebeam y Shinkfield, 1995). Entre estas perspectivas figuran la evaluación responsable (Stake, 1975), democrática (MacDonald, 1983), iluminativa (Parlett y Hamilton, 1972) y crítica (Eisner, 1981).

En la educación superior, predomina el modelo orientado a resultados, donde los estudiantes evalúan el desempeño docente a partir de su percepción. Este enfoque, aunque útil, tiende a omitir factores contextuales y procesuales, reduciendo la evaluación a una apreciación conductual limitada (Gómez y Valdez, 2019; Díaz-Barriga, 1987).

El desempeño docente integra cualidades personales y competencias profesionales esenciales para la calidad educativa. Evaluarlo implica medir la capacidad del docente para guiar y valorar el aprendizaje con eficacia y calidad (Martínez Chairez *et al.*, 2016; Carlos, 2016). Además, los docentes cumplen el papel de facilitadores y mediadores del aprendizaje (Forero Arango, 2022). La calidad educativa depende, en gran medida, de este proceso evaluativo, entendido como un esfuerzo continuo de mejora y compromiso ético de todos los actores (Vidal y Rivedero Godoy, 2023).

### **Metodología**

La validación del instrumento se enmarca en un diseño de investigación no experimental, dentro de un estudio de tipo metodológico. Este enfoque se orienta al desarrollo y evaluación de instrumentos de medición con el propósito de garantizar su validez y fiabilidad (Polit y Beck, 2008).

### **Población**

Para el desarrollo del instrumento se tomaron como

referencia diversos cuestionarios existentes para evaluar la percepción estudiantil sobre el desempeño docente, incluyendo los utilizados por la FACEN y por otras facultades de la UNA, como Ciencias Veterinarias, Ingeniería y Ciencias Agrarias. Estos instrumentos constituyeron la base de análisis y sirvieron como población de estudio para el diseño del nuevo cuestionario.

La validación del instrumento involucró dos grupos de población. En primer lugar, una muestra no probabilística de estudiantes matriculados en las distintas carreras de la Facultad, seleccionada según Hernández *et al.* (2014). Esta muestra permitió aplicar los análisis estadísticos necesarios para evaluar la consistencia interna del instrumento, garantizando la fiabilidad de los resultados dentro de las limitaciones propias del muestreo no probabilístico (Kerlinger y Lee, 2002). En segundo lugar, participaron profesionales con experiencia en evaluación educativa y desempeño docente, quienes analizaron la coherencia, relevancia y claridad de los ítems. Sus juicios aportaron una validación técnica y externa, asegurando que los ítems representen adecuadamente las dimensiones del desempeño docente.

### **Juicio de expertos**

Para verificar la coherencia, relevancia y claridad de los ítems del instrumento, se recurrió al juicio de expertos, al seguir la metodología propuesta por Escobar-Pérez y Cuervo-Martínez (2008). Conforme a los lineamientos de Grant y Davis (1997) y Hyrkäs *et al.* (2003), participaron cinco especialistas en evaluación educativa. Cada experto evaluó los ítems en función de su coherencia con los objetivos de la evaluación, su relevancia para el contexto universitario y su claridad en la redacción, utilizando una escala tipo Likert y disponiendo de un espacio para observaciones. Las sugerencias aportadas fueron analizadas y se incorporaron las modificaciones necesarias para mejorar la calidad del instrumento.

La validez de contenido se estimó a partir de las

valoraciones de los expertos mediante el coeficiente V de Aiken, aplicando la ecuación (1) y su correspondiente intervalo de confianza, calculado con las ecuaciones (2) y (3). Este coeficiente, cuyo valor varía entre 0 y 1, refleja el nivel de acuerdo entre los jueces: valores cercanos a 1 indican un alto consenso sobre la validez del ítem. El cálculo se realizó para las tres características evaluadas (relevancia, coherencia y claridad) incorporando además las observaciones de redacción en la versión final del cuestionario.

### **Consistencia interna**

Para evaluar la fiabilidad del cuestionario, se aplicó a una muestra más amplia de estudiantes. Se calculó el coeficiente de consistencia interna (alfa de Cronbach) acompañado por el cálculo de su intervalo de confianza a 95% basado en el método de Feldt (1965) para cada una de las dimensiones del cuestionario y para el instrumento en su totalidad. Un alfa de Cronbach superior a 0.7 se consideró indicativo de una buena consistencia interna.

### **Validez de constructo**

Para evaluar la validez de constructo del instrumento, se aplicó un AFC utilizando el paquete lavaan (Rosseel, 2012) del software estadístico R (R Core Team, 2022). Se definió un modelo teórico conforme lo descrito en el contexto de la investigación. Los estudiantes completaron el cuestionario proporcionando los datos necesarios para el análisis.

Se evaluó el ajuste del modelo mediante los índices de ajuste CFI, TLI y RMSEA, utilizando los siguientes criterios (Xia y Yang, 2019):

- Comparative Fit Index (CFI): un valor superior a 0.90 indica un buen ajuste.
- Tucker-Lewis Index (TLI): un valor superior a 0.90 sugiere un buen ajuste.
- Root Mean Square Error of Approximation (RMSEA): un valor inferior a 0.08 indica un ajuste aceptable.



### Procedimiento de aplicación del instrumento

Con el fin de ilustrar la aplicabilidad práctica del cuestionario validado, se procedió a su utilización en la evaluación del desempeño docente de la FACEN-UNA correspondiente al segundo semestre de 2024. La administración del instrumento estuvo a cargo de la Dirección de Talento Humano. Debido a la naturaleza sensible de los datos, los resultados se presentaron de forma anónima, identificando a los docentes como Profesor/a 1 a Profesor/a 5.

El propósito de esta aplicación fue demostrar cómo el cuestionario podía integrarse en los procesos institucionales de evaluación docente y cómo los resultados podían comunicarse tanto a nivel individual como colectivo. En el primer caso, cada docente recibió un informe personalizado con sus resultados; en el segundo, se elaboraron informes comparativos por carrera o departamento académico.

Cada informe individual incluyó una tabla con los índices obtenidos por ítem y las medias por dimensión (Responsabilidad, Didáctica, Evaluación y Promoción del Aprendizaje). Las respuestas en escala Likert de cinco categorías (1 a 5) se transformaron a una escala de 0 a 100 puntos mediante la siguiente expresión:

$$\text{Índice}_{0-100} = \frac{(\text{respuesta likert} - 1)}{4} \times 100 \quad (4)$$

De este modo, las puntuaciones pudieron interpretarse como porcentajes de logro, facilitando su comprensión. Además, se incluyeron los índices consolidados por dimensión y el índice global del docente. Los resultados se representaron mediante un gráfico tipo radar que mostró el perfil de desempeño en las cuatro dimensiones evaluadas, así como la posición del docente en los cuartiles del conjunto de su carrera o departamento, permitiendo una

comparación contextual y una interpretación integral de los resultados.

### Resultados y discusión

#### Validación del instrumento

##### Validación de contenido

La tabla 1 presenta los coeficientes de validez de contenido para la pertinencia, coherencia y claridad de cada ítem, junto con sus intervalos de confianza a 95%. La mayoría de los ítems registran valores de 1.00, lo que refleja un alto consenso entre los expertos y respalda su inclusión en la versión final del instrumento.

Los ítems 1 al 6, 9, 16 al 21 y 23 mostraron coherencia perfecta, mientras que los ítems 7, 8, 15, 18, 22, 26 y 27 presentaron valores más bajos (0.60-0.80) y mayor variabilidad, por lo que se consideró su revisión o exclusión.

En cuanto a la claridad, los ítems 1, 2, 5, 6, 9, 10, 11, 13, 21, 22, 26 y 27 alcanzaron altos coeficientes (0.93-1.00), mientras que los ítems 7, 8, 15, 18 y 24 mostraron valores menores, indicando posibles problemas de redacción o ambigüedad.

Considerando las valoraciones sobre pertinencia, coherencia y claridad, por parte de los expertos, se eliminaron los ítems 7, 8, 15, 18 y 24, teniendo como resultado final el instrumento expuesto en la tabla 2.

Agrazal-García *et al.* (2022) validaron un cuestionario para el hantavirus en Panamá mediante juicio de expertos y la V de Aiken, con valores superiores a 0.70. De modo análogo, en este estudio se aplicó la versión modificada de Aiken (Penfield y Giacobbi, 2004) para evaluar la pertinencia, coherencia y claridad de los ítems.

Tapullima-Mori *et al.* (2023) y Corrales-Álvarez *et al.* (2024) también emplearon la V de Aiken en la validación de instrumentos, confirmando su amplia aceptación y la importancia de la validez de contenido en estudios recientes.

**Tabla 1. Validez de contenido para la relevancia, coherencia y claridad de los ítems (intervalos de confianza de 95%)**

Ítem	Relevancia	Coherencia	Claridad
1	1.00 (0.80-1.00)	0.93 (0.70-0.99)	1.00 (0.80-1.00)
2	1.00 (0.80-1.00)	0.93 (0.70-0.99)	1.00 (0.80-1.00)
3	1.00 (0.80-1.00)	0.87 (0.62-0.96)	1.00 (0.80-1.00)
4	1.00 (0.80-1.00)	0.87 (0.62-0.96)	1.00 (0.80-1.00)
5	1.00 (0.80-1.00)	0.93 (0.70-0.99)	1.00 (0.80-1.00)
6	1.00 (0.80-1.00)	0.93 (0.70-0.99)	1.00 (0.80-1.00)
7	0.47 (0.25-0.70)	0.67 (0.42-0.85)	0.80 (0.55-0.93)
8	0.60 (0.36-0.80)	0.47 (0.25-0.70)	0.60 (0.36-0.80)
9	1.00 (0.80-1.00)	0.93 (0.70-0.99)	1.00 (0.80-1.00)
10	0.87 (0.62-0.96)	0.93 (0.70-0.99)	0.87 (0.62-0.96)
11	0.87 (0.62-0.96)	1.00 (0.80-1.00)	0.87 (0.62-0.96)
12	0.87 (0.62-0.96)	0.87 (0.62-0.96)	0.87 (0.62-0.96)
13	0.87 (0.62-0.96)	1.00 (0.80-1.00)	0.87 (0.62-0.96)
14	0.87 (0.62-0.96)	0.87 (0.62-0.96)	0.87 (0.62-0.96)
15	0.53 (0.30-0.75)	0.53 (0.30-0.75)	0.67 (0.42-0.85)
16	1.00 (0.80-1.00)	0.80 (0.55-0.93)	1.00 (0.80-1.00)
17	1.00 (0.80-1.00)	0.87 (0.62-0.96)	1.00 (0.80-1.00)
18	0.47 (0.25-0.70)	0.73 (0.48-0.89)	0.73 (0.48-0.89)
19	1.00 (0.80-1.00)	0.87 (0.62-0.96)	1.00 (0.80-1.00)
20	1.00 (0.80-1.00)	0.73 (0.48-0.89)	1.00 (0.80-1.00)
21	1.00 (0.80-1.00)	0.93 (0.70-0.99)	1.00 (0.80-1.00)
22	0.87 (0.62-0.96)	0.93 (0.70-0.99)	0.67 (0.42-0.85)
23	1.00 (0.80-1.00)	0.87 (0.62-0.96)	0.87 (0.62-0.96)
24	0.60 (0.36-0.80)	0.47 (0.25-0.70)	0.80 (0.55-0.93)
25	1.00 (0.80-1.00)	0.93 (0.70-0.99)	0.80 (0.55-0.93)
26	1.00 (0.80-1.00)	0.93 (0.70-0.99)	0.80 (0.55-0.93)

Fuente: elaboración propia basado en las respuestas de los expertos.



**Tabla 2. Instrumento final ajustado según sugerencias de expertos**

<b>Responsabilidad</b>
1. El docente cumplió con el horario de clase establecido (inicio y finalización).
2. El docente asistió a todas las clases planificadas en el calendario académico.
3. El docente presentó en tiempo y forma los resultados de las calificaciones de evaluaciones parciales y/o finales.
4. El docente informó con claridad sobre el programa de la asignatura.
5. El docente comunicó con claridad sobre los criterios y forma de evaluación al inicio del periodo académico.
6. El docente desarrolló todo el contenido descrito en el programa de su asignatura.
<b>Didáctica</b>
7. El docente utilizó tecnologías de la información y la comunicación adecuadas en las actividades desarrolladas en el aula (celular, aula virtual, sitios web, planillas electrónicas y otros similares).
8. El docente respondió a las consultas con paciencia y en todas las ocasiones que necesité.
9. El docente respondió con claridad y precisión a las consultas realizadas.
10. El docente desarrolló los diferentes contenidos de la asignatura de manera equilibrada (teoría y práctica).
11. El docente demostró conocimiento cabal de la asignatura.
12. El docente explicó con claridad los diferentes contenidos usando ejemplos concretos y destacando temas importantes.
13. El docente utilizó metodologías pertinentes que facilitaron el aprendizaje.
<b>Evaluación</b>
14. El nivel de las evaluaciones se ajusta de forma coherente al desarrollo de las clases.
15. El tiempo asignado para las evaluaciones parciales y finales fue coherente con el nivel de exigencia del examen.
16. Los resultados de las evaluaciones parciales fueron utilizados para retroalimentar temas de bajo rendimiento.
<b>Promoción del Aprendizaje</b>
17. El docente propició la participación de los estudiantes en clase, mediante estrategias como lluvia de ideas, preguntas, planteamiento de casos.
18. El docente promovió el respeto, la armonía y la disciplina entre compañeros.
19. El docente recibió con apertura la opinión de los estudiantes en relación con el aprendizaje propuesto en clase.
20. El docente promovió el interés por la asignatura.
21. El docente mantuvo una actitud respetuosa hacia los estudiantes.

Fuente: elaboración propia.

### **Consistencia interna**

Para evaluar la consistencia interna del instrumento, se aplicó a tres profesores de distintas asignaturas con 38, 45 y 33 estudiantes, respectivamente. Esta diversidad permitió analizar la estabilidad del instrumento en diferentes contextos docentes.

La tabla 3 presenta los coeficientes alfa de

Cronbach e intervalos de confianza a 95% para las dimensiones evaluadas. En todos los casos los valores superaron 0.90, evidenciando alta consistencia interna. Aunque la dimensión Promoción del Aprendizaje mostró valores ligeramente inferiores, todos se mantuvieron dentro de rangos óptimos para instrumentos educativos.

**Tabla 3. Consistencia interna del instrumento por dimensión, según profesor**

Dimensión	Métricas	Profesor 1	Profesor 2	Profesor 3
Responsabilidad	Cronbach	0.94	0.99	0.99
	ic 95 %	0.91-0.97	0.98-0.99	0.98-1.00
Didáctica	Cronbach	0.94	0.99	0.99
	ic 95 %	0.91-0.97	0.99-1.00	0.98-1.00
Evaluación	Cronbach	0.92	0.96	0.98
	ic 95 %	0.87-0.96	0.91-0.98	0.96-0.99
Promoción de Aprendizaje	Cronbach	0.93	0.93	0.91
	ic 95 %	0.89-0.96	0.87-0.97	0.87-0.98

Fuente: elaboración propia.

La confiabilidad de los instrumentos de medición suele evaluarse mediante el coeficiente alfa de Cronbach, considerado el más utilizado en estudios recientes por su capacidad para reflejar la consistencia interna (Tapullima-Mori *et al.*, 2023). En este trabajo también se aplicó dicho coeficiente, siguiendo los mismos criterios para estimar la fiabilidad del instrumento.

Diversas investigaciones respaldan su utilidad en contextos educativos y psicológicos. Díaz Vera *et al.* (2021) lo emplearon para analizar el impacto de las TIC en la educación superior, mientras que Apaza Gomez *et al.* (2023) obtuvieron un alfa de 0.89 en una escala de resiliencia. De forma similar, Berumen López *et al.* (2023) validaron un cuestionario sobre implicaciones educativas de la pandemia, confirmando su alta consistencia interna.

### Validez de constructo

El análisis factorial confirmatorio (AFC) evidenció un ajuste aceptable del modelo, con algunos aspectos por mejorar. El CFI fue de 0.909, superando el umbral de 0.90 (Byrne, 2012), mientras que el TLI alcanzó 0.875 y el RMSEA 0.095, valores que sugieren posibles ajustes para optimizar la estructura (Kline, 2015; Brown, 2015).

Como señalan Xia y Yang (2019), muchos modelos requieren revisión crítica por reportes incompletos o interpretaciones limitadas. En este estudio se recalca la importancia de informar adecuadamente los índices y aplicar criterios acordes al modelo. Pese a las oportunidades de mejora, la varianza explicada del 80.31 % confirma la solidez del modelo propuesto.



Los resultados de la dimensión Responsabilidad coinciden con los de Lauermann y Karabenick (2013), quienes propusieron un modelo que integra motivación, logro académico y relaciones pedagógicas. En este estudio, esta dimensión también se entiende como un compromiso ético y técnico que incide directamente en los resultados educativos (Marek, 2019).

Las dimensiones identificadas guardan correspondencia con las de la Benemérita Universidad Autónoma de Puebla (Cortés *et al.*, 2014): Evaluación se

asocia con “Evaluación de Aprendizajes”, Didáctica con “Estrategias de Enseñanza” y “Planeación”, Promoción del Aprendizaje con “Mediación” y “Relación”, y Responsabilidad con los “Ejes Transversales” del modelo Minerva.

Si bien la percepción estudiantil es una fuente válida para evaluar el desempeño docente, puede estar sujeta a sesgos de simpatía, exigencia o género (García Garduño, 2000; Mujeres con ciencia, 2019). Por ello, se recomienda complementarla con autoevaluaciones, evaluación entre pares y observación externa.

**Tabla 4. Índices del Análisis Factorial Confirmatorio**

Índices	Valor
Comparative Fit Index (CFI)	0.909
Tucker-Lewis Index (TLI)	0.875
Root Mean Square Error of Approximation (RMSEA)	0.09
Varianza total explicada	80,31%

Fuente: elaboración propia.

### Aplicación del instrumento

Los resultados del Profesor/a 1 (tabla 5) muestran un puntaje global de 4.10 en escala Likert, equivalente a un índice de 77.58 sobre 100, es decir un nivel de desempeño alto con margen de mejora.

Por dimensiones, Responsabilidad presentó el promedio más alto (4.21; 80.28), destacando la asistencia y el cumplimiento del programa. Didáctica obtuvo 4.07 (76.67), con buen uso de tecnologías y atención a consultas, aunque menor claridad en las explicaciones.

Evaluación alcanzó 4.04 (76.11), con adecuadas prácticas evaluativas, pero retroalimentación limitada. Promoción del Aprendizaje registró 4.06 (76.50), con ambiente positivo y apertura al diálogo, aunque menor motivación hacia la asignatura.

En conjunto, el desempeño del docente fue equilibrado, con fortalezas en la Responsabilidad y áreas de mejora en la Didáctica. La figura asociada presenta el perfil general mediante un gráfico radar, que permite identificar visualmente los contrastes entre dimensiones.

**Tabla 5. Percepción estudiantil sobre el desempeño docente del profesor/a 1**

<b>Criterio</b>	<b>Promedio Likert</b>	<b>Índice</b>
<b>Puntaje global</b>	<b>4.10</b>	<b>77.58</b>
<b>Dimensión 1: Responsabilidad</b>	<b>4.21</b>	<b>80.28</b>
1. El docente cumplió con el horario de clase establecido (inicio y finalización).	4.17	79.17
2. El docente asistió a todas las clases planificadas en el calendario académico.	4.33	83.33
3. El docente presentó en tiempo y forma los resultados de las calificaciones de evaluaciones parciales y/o finales.	4.17	79.17
4. El docente informó con claridad sobre el programa de la asignatura.	4.20	80.00
5. El docente comunicó con claridad sobre los criterios y forma de evaluación al inicio del periodo académico.	4.13	78.33
6. El docente desarrolló todo el contenido descrito en el programa de su asignatura.	4.27	81.67
<b>Dimensión 2: Didáctica</b>	<b>4.07</b>	<b>76.67</b>
7. El docente utilizó tecnologías de la información y la comunicación adecuadas en las actividades desarrolladas en el aula (celular, aula virtual, sitios web, planillas electrónicas y otros similares).	4.20	80.00
8. El docente respondió a las consultas con paciencia y en todas las ocasiones que necesité.	4.17	79.17
9. El docente respondió con claridad y precisión a las consultas realizadas.	4.23	80.83
10. El docente desarrolló los diferentes contenidos de la asignatura de manera equilibrada (teoría y práctica).	4.00	75.00
11. El docente demostró conocimiento cabal de la asignatura.	3.90	72.50
12. El docente explicó con claridad los diferentes contenidos usando ejemplos concretos y destacando temas importantes.	3.97	74.17
13. El docente utilizó metodologías pertinentes que facilitaron el aprendizaje.	4.00	75.00
<b>Dimensión 3: Evaluación</b>	<b>4.04</b>	<b>76.11</b>
14. El nivel de las evaluaciones se ajusta de forma coherente al desarrollo de las clases.	4.20	80.00
15. El tiempo asignado para las evaluaciones parciales y finales fue coherente con el nivel de exigencia del examen.	3.97	74.17
16. Los resultados de las evaluaciones parciales fueron utilizados para retroalimentar temas de bajo rendimiento.	3.97	74.17
<b>Dimensión 4: Promoción del Aprendizaje</b>	<b>4.06</b>	<b>76.50</b>
17. El docente propició la participación de los estudiantes en clase, mediante estrategias como lluvia de ideas, preguntas, planteo de casos.	4.17	79.17
18. El docente promovió el respeto, la armonía y la disciplina entre compañeros.	3.97	74.17
19. El docente recibió con apertura la opinión de los estudiantes en relación con el aprendizaje propuesto en clase.	4.13	78.33
20. El docente promovió el interés por la asignatura.	4.07	76.67
21. El docente mantuvo una actitud respetuosa hacia los estudiantes.	3.97	74.17

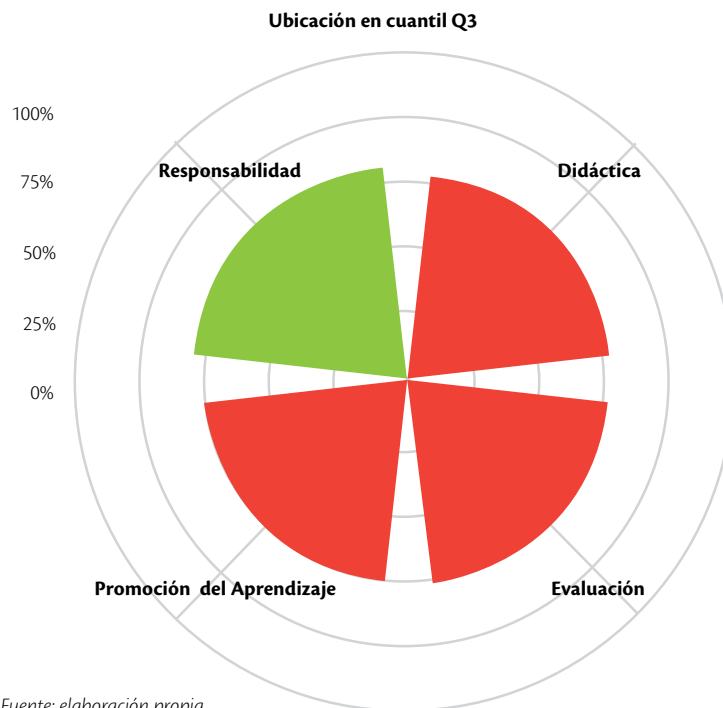
Fuente: elaboración propia.



La gráfica 1 muestra de manera gráfica el perfil del Profesor/a 1 en las cuatro dimensiones del cuestionario. Para facilitar la interpretación, se emplea un esquema de tres colores: verde para índices altos ( $\geq 80$ ), amarillo para niveles intermedios (60-79.9) y rojo para niveles bajos ( $< 60$ ). En este caso, la dimensión de Responsabilidad alcanza un nivel alto, mientras que las dimensiones de Didáctica, Evaluación y Promoción del Aprendizaje se ubican en el rango intermedio, lo cual refleja un desempeño equilibrado con posibilidades de mejora.

Además, se considera el índice global de cada profesor/a y, tomando como referencia a todos los docentes de la misma carrera o departamento académico, se determina su ubicación en uno de los cuatro cuartiles. Esta clasificación otorga un criterio ordinal que permite situar al docente en relación con sus pares. En el caso del Profesor/a 1, su índice global corresponde al tercer cuartil (Q3), lo cual indica que, aunque su desempeño no se encuentra entre los más bajos, tampoco se ubica dentro del grupo de mejor valoración en el conjunto académico de referencia.

**Gráfica 1. Perfil del Profesor/a 1 en las cuatro dimensiones del cuestionario de evaluación docente**



### Informe para la unidad académica

A continuación, se presentan los resultados agregados que permiten comparar el desempeño de varios profesores/as de una misma carrera o departamento académico. Estos resultados constituyen un insumo valioso para identificar fortalezas comunes y áreas de mejora, y facilitan la toma de decisiones orientadas a la implementación de planes de acompañamiento

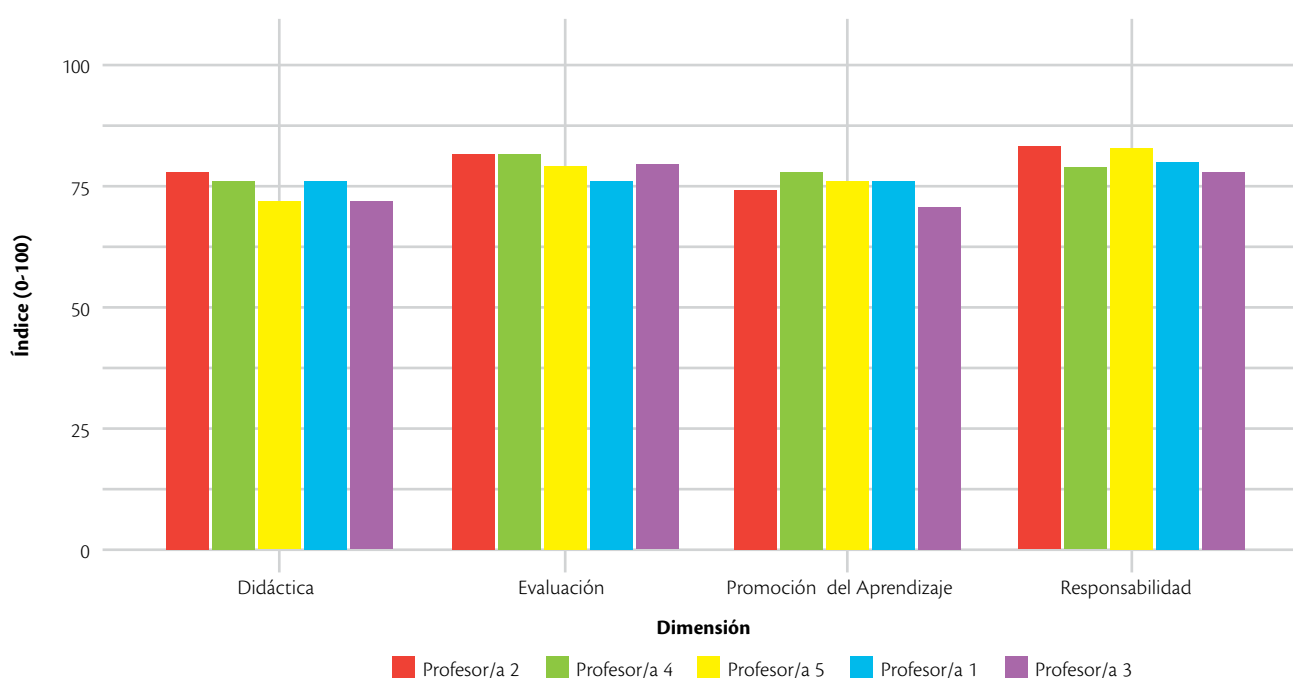
y capacitación docente. El análisis colectivo complementa los informes individuales y proporciona una visión comparativa del contexto académico en su conjunto.

La gráfica 2 presenta la comparación de los índices de desempeño docente por dimensión y profesor/a. Se observa que la Responsabilidad constituye una fortaleza común a todo el grupo de docentes, con

puntajes elevados y escasa variabilidad, lo que refleja un cumplimiento uniforme de los compromisos académicos básicos. Por el contrario, la Didáctica aparece como la dimensión con resultados relativamente más bajos y mayor dispersión, lo que la convierte en una debilidad compartida que requiere

atención prioritaria. Las dimensiones de Evaluación y Promoción del Aprendizaje se sitúan en valores intermedios, mostrando un desempeño aceptable en conjunto, aunque con posibilidades de mejora en la retroalimentación y en las estrategias para fomentar la motivación estudiantil.

**Gráfica 2. Comparación de los índices de desempeño docente por dimensión y profesor/a**



Fuente: elaboración propia.

A partir de los resultados obtenidos para los cinco profesores/as analizados, la institución puede orientar planes de mejora que atiendan tanto a las fortalezas como a las debilidades detectadas. El cumplimiento de la responsabilidad académica constituye un aspecto consolidado que debe mantenerse como estándar institucional. En cambio, la dimensión didáctica muestra mayor margen de mejora, lo cual sugiere implementar programas de capacitación en estrategias de enseñanza activa, uso de recursos pedagógicos y claridad expositiva. Los resultados de la dimensión evaluación apuntan a fortalecer la

retroalimentación, mientras que, en promoción del aprendizaje, conviene potenciar acciones que estimulen la motivación y participación estudiantil. Así, la aplicación del instrumento permite diagnosticar el desempeño docente y generar insumos concretos para políticas de formación y acompañamiento en la FACEN.

En la interpretación de los resultados, un aspecto relevante es la definición de categorías de desempeño. El uso de umbrales absolutos en la escala (por ejemplo, 0-60 “mucho por mejorar”, 61-70 “algo por mejorar”, 71-80 “poco por mejorar”, 81-90



“buen desempeño” y 91-100 “excelente”) presenta limitaciones, ya que pequeñas variaciones en las valoraciones no siempre implican cambios en la clasificación y se pierde sensibilidad en la medición. Además, estos umbrales fijos no consideran las diferencias entre contextos o disciplinas, lo que puede generar clasificaciones poco equitativas entre facultades (Uttl y Smibert, 2017).

Debido a que las evaluaciones docentes suelen concentrarse en valores altos (efecto techo), un mismo umbral absoluto puede sobrestimar o subestimar el desempeño según el contexto. Por ello, resulta más apropiado emplear medidas de posición relativas, como cuartiles o percentiles, que sitúan a cada profesor/a en relación con sus pares dentro de la carrera o el departamento. Este enfoque incrementa la sensibilidad del análisis y permite una interpretación más contextualizada y justa (Pontificia Universidad Católica del Perú [PUCP], 2022).

El uso de medidas de posición relativas mejora la capacidad de discriminación del instrumento y favorece una interpretación más equitativa del desempeño docente dentro del entorno institucional.

### **Seguimiento en el tiempo**

El uso de los índices de desempeño a lo largo de distintos periodos permite construir series temporales para cada profesor/a. Este seguimiento posibilita identificar tanto las mejoras sostenidas en la percepción estudiantil como las desviaciones que puedan reflejar descensos en el desempeño, ya sea en el puntaje global o en dimensiones específicas. De esta manera, la información obtenida trasciende el carácter diagnóstico de un momento particular y se convierte en un insumo para evaluar la evolución del quehacer docente. La unidad académica responsable de diseñar las estrategias de mejora debe monitorear de manera sistemática el impacto de los planes implementados, de modo que el instrumento cumpla efectivamente con su propósito de orientar la retroalimentación y el fortalecimiento continuo de la docencia.

### **Conclusiones y recomendaciones**

El instrumento validado consta de cuatro dimensiones clave que abarcan aspectos fundamentales del desempeño docente:

- **Responsabilidad:** compuesta por seis ítems que evalúan la puntualidad, asistencia y responsabilidad del docente en la preparación y entrega de contenidos.
- **Didáctica:** con siete ítems, esta dimensión mide la efectividad de las metodologías de enseñanza, la claridad en la explicación de contenidos y la capacidad de mantener el interés de los estudiantes.
- **Evaluación:** consta de tres ítems que valoran la coherencia entre las evaluaciones y los contenidos impartidos, la claridad en los criterios de evaluación y la retroalimentación proporcionada a los estudiantes.
- **Promoción del Aprendizaje:** incluye cinco ítems que miden la capacidad del docente para fomentar un ambiente de respeto y participación, además de motivar y apoyar a los estudiantes en su proceso de aprendizaje.

El cuestionario diseñado para evaluar el desempeño docente en la Facultad de Ciencias Exactas y Naturales demuestra una alta consistencia interna en todas las dimensiones, con coeficientes de Cronbach superiores a 0.90 en la mayoría de los casos. Estos resultados indican que los ítems son coherentes y miden constructos similares de manera consistente.

El uso del índice V de Aiken, ajustado según la metodología de Penfield y Giacobbi (2004), permitió evaluar la pertinencia, coherencia y claridad de cada ítem del cuestionario, lo que resultó clave para afinar el instrumento hasta alcanzar su versión final. Este proceso aseguró la validez de contenido del cuestionario, asegurando que las dimensiones evaluadas reflejaran de manera precisa y efectiva el constructo del desempeño docente. Los ítems seleccionados para la versión final mostraron altos

coeficientes de validez, superiores a 0.87, lo cual refleja un elevado grado de acuerdo entre los expertos en cuanto a la relevancia, coherencia y claridad de los ítems.

El AFC validó la estructura teórica del cuestionario, compuesta por las cuatro dimensiones propuestas. Los índices de ajuste obtenidos (CFI, TLI y RMSEA) demostraron un ajuste adecuado del modelo, lo que confirma la idoneidad de la estructura teórica para medir con eficacia el desempeño docente en las dimensiones definidas.

La aplicación práctica del instrumento durante la evaluación docente del segundo semestre de 2024 en la FACEN permitió ilustrar cómo los resultados individuales y colectivos ofrecen información útil para la toma de decisiones. En los informes individuales, cada docente recibe sus índices por ítem y por dimensión, junto con un índice global y su ubicación en los cuartiles de referencia. En los informes colectivos, los resultados permiten identificar fortalezas comunes, como el cumplimiento de la responsabilidad académica, y debilidades compartidas, como las variaciones en la dimensión didáctica. Este ejemplo demuestra que el instrumento no solo es válido y confiable, sino también aplicable y relevante en contextos reales de gestión académica.

Este estudio adquiere particular relevancia en el contexto de América Latina, donde el desarrollo de la calidad educativa continúa constituyendo un desafío en las instituciones de educación superior. La validación del uso de la percepción estudiantil como una fuente clave de retroalimentación reafirma su valor como una herramienta pertinente para evaluar el desempeño docente. Este enfoque no sólo contribuye a elevar los estándares de enseñanza y

aprendizaje, sino que también refuerza la importancia de la evaluación docente como un pilar fundamental para apuntalar la calidad educativa.

Se recomienda expandir el uso de este instrumento a otras instituciones de educación superior tanto en Paraguay como América Latina. Su aplicación contribuiría a obtener datos consistentes y comparables, facilitando la identificación de áreas de mejora y apoyando decisiones informadas para el desarrollo profesional docente, fomentando así la elevación de los estándares educativos.

Es fundamental realizar una revisión y actualización periódica del cuestionario para asegurar su relevancia y coherencia con las prácticas educativas en constante evolución y las expectativas de los estudiantes. En América Latina, donde los sistemas educativos enfrentan transformaciones continuas, la incorporación de nuevas dimensiones o ítems, en respuesta a las necesidades emergentes, puede fortalecer la validez y efectividad del instrumento, adaptándolo mejor a los desafíos específicos del contexto educativo.

Futuros estudios deberían enfocarse en la validación del cuestionario en distintos contextos y poblaciones de América Latina. Asimismo, se recomienda el análisis y aplicación de otros métodos de evaluación del desempeño docente que complementen la perspectiva estudiantil con enfoques diversos y multidimensionales, tales como la autoevaluación, la evaluación entre pares y la observación externa. La combinación de estos enfoques permitiría obtener una visión más integral y equilibrada del desempeño docente, lo que contribuiría al desarrollo de políticas educativas orientadas para fortalecer la calidad en la educación superior. ■



## Referencias

- Agrazal-García, Janeth, Lydia Gordon-de Isaacs y Ricaurte Tuñón (2022), "Design and validation of a questionnaire to measure hantavirus preventive practices in an endemic community", *Revista Peruana de Medicina Experimental y Salud Pública*, vol. 39, núm. 1, pp. 47-54, DOI: <<https://doi.org/10.17843/rpmesp.2022.391.9740>>.
- Aiken, Lewis R. (1980), "Content validity and reliability of single items or questionnaires", *Educational and Psychological Measurement*, vol. 40, núm. 4, pp. 955-959, DOI: <<https://doi.org/10.1177/001316448004000419>>.
- Aiken, Lewis R. (1985), "Three coefficients for analyzing the reliability and validity of ratings", *Educational and Psychological Measurement*, vol. 45, núm. 1, pp. 131-142, DOI: <<https://doi.org/10.1177/0013164485451012>>.
- Anastasi, Anne y Susana Urbina (1998), *Test psicológicos*, México, Pearson Educación.
- Apaza Gomez, Fanny, Verónica Villanera Zelaya, Dafne Marleny Alarcon Mamani, María Emma Zúñiga Vásquez y Yuselino Maquera Maquera (2023), "Resiliencia en adolescentes expuestos a contextos de riesgo psicosociales: una revisión sistemática", *Ciencia Latina Revista Científica Multidisciplinar*, vol. 7, núm. 1, pp. 5434-5445, DOI: <[https://doi.org/10.37811/cl\\_rcm.v7i1.4839](https://doi.org/10.37811/cl_rcm.v7i1.4839)>.
- Bartolomé, Margarita (1983), *Validez de los instrumentos empleados en los procesos de medida. Evaluación y assessment, Ponencia inédita*, Barcelona, Universidad de Barcelona.
- Berumen López, Efrén, Héctor Gabriel Villegas Berumen y Susana Ávila Acevedo (2023), "Implicaciones de la educación virtual durante la pandemia COVID-19: una encuesta a estudiantes del Tecnológico Nacional de México", *RIDE. Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, vol. 13, núm. 26, e445, DOI: <<https://doi.org/10.23913/ride.v13i26.1404>>.
- Brown, Timothy Anthony (2015), *Confirmatory factor analysis for applied research*, Nueva York, Guilford Press, DOI: <<https://doi.org/10.4324/9780203807644>>.
- Byrne, Barbara Marie (2012), *Structural equation modeling with mplus: basic concepts, applications, and programming*, Nueva York, Routledge, DOI: <<https://doi.org/10.4324/9780203807644>>.
- Cappello, Héctor Manuel, María Isabel Aguirre, Rosa Delia Castro, Rosa Delia Cervantes, Juana Imelda Infante y Raúl Marín (2004), "Estudio del análisis de la consistencia interna y la validez de construcción de una prueba en el programa de enseñanza vivencial de las ciencias", *Revista Internacional de Ciencias Sociales y Humanidades SOCIOTAM*, vol. 14, núm. 2, pp. 47-72. <https://www.redalyc.org/articulo.oa?id=65414202>>.
- Carlos Guzmán, Jesús (2016), "¿Qué y cómo evaluar el desempeño docente? Una propuesta basada en los factores que favorecen el aprendizaje", *Propósitos y Representaciones*, vol. 4, núm. 2, pp. 285-358, DOI: <<https://doi.org/10.20511/pyr2016.v4n2.124>>.
- Carmines, Edward Gordon y Richard Alan Zeller (1979), *Reliability and Validity Assesment*, Beverly Hills, Sage Publications.
- Corrales-Álvarez, Milena, Lina Marcela Ocampo y Sergio Augusto Cardona Torres (2024), "Instrumentos de evaluación del pensamiento computacional: una revisión sistemática", *TecnoLógicas*, vol. 27, núm. 59, e2950, DOI: <<https://doi.org/10.22430/22565337.2950>>.
- Cortés, Ernesto, Margarita Campos y María P. Moreno (2014), "Priorización de las dimensiones de evaluación al desempeño docente por el estudiante, en tres áreas del conocimiento", *Formación Universitaria*, vol. 7, núm. 2, pp. 3-10, DOI: <<http://dx.doi.org/10.4067/S0718-50062014000200002>>.
- Cronbach, Lee y Paul Meehl (1955), "Construct validity in psychological tests", *Psychological Bulletin*, vol. 52, núm. 4, pp. 281-302, DOI: <<https://doi.org/10.1037/h0040957>>.
- De Chaparro, Gladys Jaime, Luz Janneth Romero Farfán, Edilberto Rincón C. y Luis H. Jaime G. (2008), "Evaluación de desempeño docente", *Cuadernos de Lingüística Hispánica*, núm. 11, pp. 167-178, <<https://www.redalyc.org/articulo.oa?id=322227496013>> [Consulta: agosto de 2024].
- Díaz-Barriga, Frida (1987), "El pensamiento del

- adolescente y el diseño curricular en educación media superior”, *Perfiles Educativos*, núm. 37, pp. 16-26.
- Díaz Vera, Janeth Pilar, Alicia Karina Ruiz Ramírez y Carolina Egúez Cevallos (2021), “Impacto de las TIC: desafíos y oportunidades de la Educación Superior frente al COVID-19”, *Revista Científica UISRAEL*, vol. 8, núm. 2, pp. 113-134, DOI: <<https://doi.org/10.35290/rcui.v8n2.2021.448>>.
- Divoky, James J. y Mary A. Rothermel (1988), “Student perceptions of the relative importance of dimensions of teaching performance across type of class”, *Educational Research Quarterly*, vol. 12, núm. 3, pp. 40-45, <<https://psycnet.apa.org/record/1990-15794-001>> [Consulta: agosto de 2024].
- Eisner, Elliot Wayne (1981), “On the differences between scientific and artistic approaches to qualitative research”, *Educational Researcher*, vol. 10, núm. 4, pp. 5-9, DOI: <<https://doi.org/10.3102/0013189X010004005>>.
- Escobar-Pérez, Jazmine y Ángela Cuervo-Martínez (2008), “Validez de contenido y juicio de expertos: una aproximación a su utilización”, *Avances en Medición*, vol. 6, núm. 1, pp. 27-36, <[https://www.humanas.unal.edu.co/lab\\_psicometria/application/files/9416/0463/3548/Vol\\_6\\_Articulo3\\_Juicio\\_de\\_expertos\\_27-36.pdf](https://www.humanas.unal.edu.co/lab_psicometria/application/files/9416/0463/3548/Vol_6_Articulo3_Juicio_de_expertos_27-36.pdf)> [Consulta: diciembre de 2023].
- Escribano Hervis, Elmys (2018), “El desempeño del docente como factor asociado a la calidad educativa en América Latina”, *Revista Educación*, vol. 42, núm. 2, pp. 738-752, DOI: <<https://doi.org/10.15517/revedu.v42i2.27033>>.
- Feldt, Leonard Samuel (1965), “The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty”, *Psychometrika*, vol. 30, núm. 3, pp. 357-370, DOI: <<https://doi.org/10.1007/BF02289499>>.
- Forero Arango, Ximena (2022), “El papel de la interacción en la educación superior: hacia modelos pedagógicos más flexibles”, *EduTec. Revista Electrónica de Tecnología Educativa*, núm. 79, pp. 134-148, DOI: <<https://doi.org/10.21556/edutec.2022.79.2363>>.
- García Garduño, José María (2000), “¿Qué factores extraclase o sesgos afectan la evaluación docente en la educación superior?”, *Revista Mexicana de Investigación Educativa*, vol. 5, núm. 10, pp. 303-325, <<https://dialnet.unirioja.es/servlet/articulo?codigo=299841>> [consulta: febrero 2024].
- Gómez, Luis Fernando y María Guadalupe Valdés (2019), “La evaluación del desempeño docente en la educación superior”, *Propósitos y Representaciones*, vol. 7, núm. 2, pp. 479-515, DOI: <<https://doi.org/10.20511/pyr2019.v7n2.255>>.
- Gonzales, Roberto (2022), “Desempeño docente y logro de aprendizajes en estudiantes universitarios”, *Revista Innova Educación*, vol. 4, núm. 2, pp. 25-44, DOI: <<https://doi.org/10.35622/j.rie.2022.02.002>>.
- Grant, Joan Suzanne y Linda Louise Davis (1997), “Selection and use of content experts for instrument development”, *Research in Nursing & Health*, vol. 20, núm. 3, pp. 269-274, DOI: <[https://doi.org/10.1002/\(sici\)1098-240x\(199706\)20:3%3C269::aid-nur9%3E3.0.co;2-g](https://doi.org/10.1002/(sici)1098-240x(199706)20:3%3C269::aid-nur9%3E3.0.co;2-g)>.
- Henríquez Ritchie, Patricio, Juan Carlos Pérez Morán, Sofía Contreras Roldán y Carlos Javier del Cid (2024), “Evaluación del desempeño docente por estudiantes de una universidad pública del noroeste de México: análisis descriptivos y comparativos”, *RIDE Revista Iberoamericana para la Investigación y El Desarrollo Educativo*, vol. 15, núm. 29, e738, DOI: <<https://doi.org/10.23913/ride.v15i29.2100>>.
- Hernández, Héctor Andrés y Alina Eugenia Pascual Barrera (2018), “Validación de un instrumento de investigación para el diseño de una metodología de autoevaluación del sistema de gestión ambiental”, *Revista de Investigación Agraria y Ambiental*, vol. 9, núm. 1, pp. 157-164, DOI: <<https://doi.org/10.22490/21456453.2186>>.
- Hernández, Roberto, Carlos Fernández y Pilar Baptista (2014), *Metodología de la investigación*, México, McGraw-Hill.
- Hoyle, Rick H. (2012), *Handbook of structural equation modeling*, Nueva York, Guilford Press.
- Hyrkäs, Kristiina, Kaija Appelqvist-Schmidlechner y Lea Oksa (2003), “Validating an instrument for clinical supervision using an expert panel”, *International Journal*



- of *Nursing Studies*, vol. 40, núm. 6, pp. 619-625, DOI: <[https://doi.org/10.1016/S0020-7489\(03\)00036-1](https://doi.org/10.1016/S0020-7489(03)00036-1)>.
- Jornet Meliá, Jesús Miguel, María Jesús Perales Montolio y José González Such (2020), “El concepto de validez de los procesos de evaluación de la docencia”, *Revista Española de Pedagogía*, vol. 78, núm. 276, pp. 233-252, DOI: <<https://doi.org/10.22550/REP78-2-2020-01>>.
- Kerlinger, Fred Nathaniel y Howard Burton Lee (2002), *Investigación del comportamiento. métodos de investigación en Ciencias Sociales*, México, McGraw Hill.
- Kline, Rex Brown (2015), *Principles and practice of structural equation modeling*, Nueva York, Guilford Press.
- Lauermann, Fani y Stuart A. Karabenick (2013), “The meaning and measure of teachers’ sense of responsibility for educational outcomes”, *Teaching and Teacher Education*, vol. 30, núm. 1, pp. 13-26, DOI: <<https://doi.org/10.1016/j.tate.2012.10.001>>.
- Ley-Leyva, Nelly Victoria y Eudaldo Enrique Espinoza-Freire (2021), “Características de la evaluación educativa en el proceso de aprendizaje”, *Revista Universidad y Sociedad*, vol. 13, núm. 6, pp. 363-370, <[http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S2218-36202021000600363&lng=es&tlng=pt](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2218-36202021000600363&lng=es&tlng=pt)> [Consulta: junio de 2024].
- MacDonald, Barry (1983), “La evaluación y el control de la educación”, en José Gimeno y Ángela Pérez, *La enseñanza, su teoría y su práctica*, Madrid, Akal, pp. 467-478.
- Magnusson, David (1966), *Test Theory, Reading*, Addison-Wesley Publishing Company.
- Manzi, Jorge, María Rosa García y Sandy Taut (eds.) (2021), *Validity of educational assessments in Chile and Latin America*, Cham, Springer International Publishing, DOI: <<https://doi.org/10.1007/978-3-030-78390-7>>.
- Marek, Lidia (2019), “Responsibility as an element of the teacher’s professional competences in a scope of evaluating educational work”, *Problemy Profesjologii*, núm. 2, pp. 45-52, <<https://www.ceeol.com/search/article-detail?id=923697>> [Consulta: julio de 2024].
- Martínez Chairez, Guadalupe Iván, Albertico Guevara Araiza, María Manuela Valles Ornelas (2016), “El desempeño docente y la calidad educativa”, *Ra Ximhai: revista científica de sociedad, cultura y desarrollo sostenible*, vol. 12, núm. 6, pp. 123-134, <<https://dialnet.unirioja.es/servlet/articulo?codigo=7933112>> [Consulta: septiembre de 2023].
- McNeish, Daniel y Melissa Gordon Wolf (2020), “Thinking twice about sum scores”, *Behav Res*, vol. 52, núm. 6, pp. 2287-2305, DOI: <<https://doi.org/10.3758/s13428-020-01398-0>>.
- Mohammadi, Mohammad (2021), “Dimensions of Teacher Performance Evaluation by Students in Higher Education”, *Shanlax International Journal of Education*, vol. 9, núm. 2, pp. 18-25, <<https://eric.ed.gov/?id=EJ1287514>> [Consulta: junio de 2024].
- Mujeres con ciencia (2019), “Sesgo de género en las evaluaciones docentes”, 17 de abril, Cátedra de Cultura Científica-Universidad del País Vasco, <<https://mujeresconciencia.com/2019/04/17/sesgo-de-genero-en-las-evaluaciones-docentes/>> [Consulta: noviembre de 2023].
- Pacheco Cámara, María Luisa del Carmen, Isela Ibarra Bocardo, Miriam Elizabeth Iñiguez Galindo, Héctor Lee García, Claudia Victoria Sánchez Sánchez (2018), “La evaluación del desempeño docente en la educación superior”, *Revista digital universitaria*, vol. 19, núm. 6, pp. 1-11, DOI: <<http://doi.org/10.22201/codeic.16076079e.2018.v19n6.a2>>.
- Parlett, Malcolm y David Hamilton (1972), “Evaluation and Illumination”, reimpresso en Tawney, David (1976), *Curriculum Reevaluation Today. Trends and Implications*, London, McMillan, pp. 39-53.
- Penfield, Randall D., y Peter R. Giacobbi (2004), “Applying a score confidence interval to Aiken’s Item Content-Relevance Index”, *Measurement in Physical Education and Exercise Science*, vol. 8, núm. 4, pp. 213-225, DOI: <[https://doi.org/10.1207/s15327841mpee0804\\_3](https://doi.org/10.1207/s15327841mpee0804_3)>.
- Polit, Denise. F. y Cheryl Tatano Beck (2008), *Nursing research: Generating and assessing evidence for nursing practice*, Filadelfia, Lippincott Williams & Wilkins.
- Pontificia Universidad Católica del Perú (2022), “Nueva escala de evaluación docente para una mayor precisión en los resultados”, *PUCP*, <<https://www.pucp.edu.pe/>

- noticias/nueva-escala-de-evaluacion-docente-para-una-mayor-precision-en-los-resultados>, [Consulta: noviembre de 2023].
- R Core Team (2022), *R: a language and environment for statistical computing*, Viena, R Foundation for Statistical Computing, <<https://www.R-project.org/>> [Consulta: diciembre de 2023].
- Reynolds, Cecil R., Robert A. Altmann y Daniel N. Allen (2021), "Validity", en *Mastering Modern Psychological Testing*, Cham, Springer International Publishing, pp. 185-222, DOI: <[https://doi.org/10.1007/978-3-030-59455-8\\_5](https://doi.org/10.1007/978-3-030-59455-8_5)>.
- Rosseel, Yves (2012), "lavaan: An R Package for Structural Equation Modeling", *Journal of Statistical Software*, vol. 48, núm. 2, pp. 1-36, DOI: <<https://doi.org/10.18637/jss.v048.i02>>.
- Schumacker, Rick E. y Reginald G. Lomax (2004), *A beginner's guide to structural equation modeling*, Mahwah, Lawrence Erlbaum Associates, Inc.
- Soriano Rodríguez, Ana María, (2014), "Diseño y validación de instrumentos de medición", *Revista Diálogos*, vol. 8, núm. 14, pp. 19-40, <<https://www.revistas.udb.edu.sv/ojs/index.php/dl/article/view/166>> [Consulta: diciembre de 2023].
- Stake, Robert E. (1975), *Evaluating the art in education: a responsive approach*, Columbus, Charles E. Merrill Publishing Company.
- Stronge, James H. (2018), *Qualities of effective teachers*, Alexandria, ASCD.
- Stufflebeam, Daniel L. (2003), "The CIPP model for evaluation", en T. Kellaghan (ed.), *International handbook of educational evaluation*, Dordrecht, Kluwer Academic, pp. 31-62.
- Stufflebeam, Daniel L. y Anthony J. Shinkfield (1995), *Evaluación sistemática: guía teórica y práctica*, Barcelona, Paidós.
- Tapullima-Mori, Calixto, José Livia Segovia, Nieves del Pilar Pizzán Tomanguillo, Sandra Lucero Pizzán Tomanguillo, Milagros Inípe Cachay, Astrid Irene Saenz Chisquipama y Fiorella Gómez Sangama (2023), "Propiedades psicométricas de las escalas de competencias investigativas: una revisión sistemática", *Propósitos y Representaciones*, vol. 11, núm. 3, e1868, DOI: <<https://dx.doi.org/10.20511/pyr2023.v11n3.1868>>.
- Uljens, Michael (2004), *School didactics and learning: A school didactic model framing an analysis of pedagogical implications of learning theory*, Hove, Psychology Press Ltd.
- Uttl, Bob y Dylan Smibert (2017), "Student evaluations of teaching: Teaching quantitative courses can be hazardous to one's career", *PeerJ*, vol. 5, e3299, DOI: <<https://doi.org/10.7717/peerj.3299>>.
- Valentín-Martínez, Bilda y Cristina Mayor-Ruiz (2023), "La evaluación del profesorado universitario: una mirada del estudiantado", *Educação e Pesquisa*, vol. 49 (continuo), e241907, DOI: <<https://doi.org/10.1590/S1678-4634202349241907esp>>.
- Vidal, Melania y Matías Rivadero Godoy (2023), "La gestión del desarrollo: análisis de las principales causas que influyen negativamente en el indicador visualizo oportunidades de desarrollo y crecimiento en la empresa, de la encuesta anual de clima laboral de la empresa RIDAL SA, en el periodo 2019-2022", tesis de licenciatura en Gestión de Recursos Humanos, Argentina, Centro Universitario Regional Zona Atlántica-Universidad Nacional de Comahue, <<http://rdi.uncoma.edu.ar/handle/uncomaid/17643>> [Consulta: octubre de 2023].
- Xia, Yan y Yanyun Yang (2019), "RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods", *Behav Res*, vol. 51, pp. 409-428, DOI: <<https://doi.org/10.3758/s13428-018-1055-2>>.

#### Cómo citar este artículo:

Rivas-Martínez, Gustavo, Luis Ortiz y Fátima Bañuelos-Gómez (2026), "Diseño y validación de un instrumento para la evaluación del desempeño docente en la educación superior", *Revista Iberoamericana de Educación Superior (RIES)*, vol. XVII, núm. 48, pp. 61-81, DOI: <https://doi.org/10.22201/iisue.20072872e.2026.48.1856>.