



## Construcción y validación de la Escala de Evaluación de Innovaciones Educativas en educación superior

### **Carola Bruna Jofré\***

Universidad de Concepción  
Facultad de Ciencias Biológicas

Mail: [carolabruna@udec.cl](mailto:carolabruna@udec.cl)

ORCID: <https://orcid.org/0000-0003-0554-7147>

### **Claudio Bustos Navarrete**

Universidad de Concepción  
Facultad de Psicología

Mail: [cbustos@udec.cl](mailto:cbustos@udec.cl)

ORCID: <https://orcid.org/0000-0003-3478-9858>

### **Karla Lobos Peña**

Universidad de Concepción  
Dirección de Docencia

Mail: [karlalobos@udec.cl](mailto:karlalobos@udec.cl)

ORCID: <https://orcid.org/0000-0002-0145-0041>

### **Camila Espinoza-Parçet**

Universidad de Concepción  
Dirección de Docencia

Mail: [cespinoza@udec.cl](mailto:cespinoza@udec.cl)

ORCID: <https://orcid.org/0000-0002-7655-4881>

### **Felipe Sáez-Noriega**

University of Manchester  
Alliance Manchester Business School

Mail: [felipe.saeznoriega@postgrad.manchester.ac.uk](mailto:felipe.saeznoriega@postgrad.manchester.ac.uk)

ORCID: <https://orcid.org/0000-0003-1928-6072>

### **Carolyn Fernández Branada**

Universidad de Concepción  
Facultad de Educación

Mail: [carferna@udec.cl](mailto:carferna@udec.cl)

ORCID: <https://orcid.org/0000-0003-4262-8026>

\*Autora de correspondencia: [carolabruna@udec.cl](mailto:carolabruna@udec.cl)

Recibido: 07/04/2025 – Aceptado: 18/12/2025

ISSN: 0210-2773

DOI: <https://doi.org/10.17811/rifie.22573>



## RESUMEN

Los desafíos de la educación superior interpelan al profesorado a innovar en las prácticas docentes para promover el aprendizaje activo y el desarrollo de habilidades y actitudes. Si bien se han reportado esfuerzos del profesorado y mecanismos institucionales para abordar esta necesidad, las innovaciones son escasamente evaluadas para tomar decisiones pedagógicas basadas en evidencia. Este trabajo se enfocó en la elaboración y validación de una escala para evaluar innovaciones educativas en educación superior, de uso amplio para la diversidad de diseños para mejorar el proceso enseñanza-aprendizaje. Participaron 450 estudiantes universitarios que experimentaron una innovación en alguna asignatura. Con una mitad de la muestra, se evaluaron tres modelos mediante análisis factorial semi confirmatorio, ajustando la estructura factorial. Con la segunda, se aplicó un análisis factorial confirmatorio usando los modelos con buen ajuste de la primera submuestra. Como resultado, se identificó un modelo bifactorial, con un factor general y cuatro factores específicos. La Escala de Evaluación de Innovaciones Educativas (EVINED) es válida y confiable (alfa = 0,97; omega total = 0,98). Este instrumento es de utilidad para el profesorado y administradores de las políticas educativas institucionales, facilitando la instalación de una cultura de la evidencia en la práctica docente.

*Palabras clave:* instrumento de evaluación, validez del instrumento, innovación educativa, educación superior, evaluación psicométrica.

## Construction and validation of the Educational Innovation Assessment Scale in higher education

## ABSTRACT

The challenges of higher education require faculty to innovate in teaching practices to promote active learning and the development of skills and attitudes. While there have been reported efforts from educators and institutional support mechanisms to respond to this need, there is little evaluation of these innovations for evidence-based pedagogical decision-making. This study focused on developing and validating a scale designed to assess educational innovations in higher education, applicable across diverse designs, aimed at enhancing the teaching-learning process. An instrumental design was employed. 450 university students who had experienced an innovation in a course participated. With half of the sample, three models were evaluated through a semi-confirmatory factor analysis to refine the factor structure. With the second half, a strict confirmatory factor analysis was conducted on the models that demonstrated a good fit in the first phase. As a result, a two-factor model was identified, with one general factor and four specific factors. The Educational Innovations Evaluation Scale (EVINED) is valid and reliable (alfa = .97; total omega = .98). This tool can be useful for teachers and administrators of institutional educational policies, thereby fostering a culture of evidence-based pedagogical improvement.

*Keywords:* assessment instrument, instrument validity, educational innovation, higher education, psychometric evaluation.

### 1. Introducción

La innovación es un factor crucial para la competitividad de las organizaciones, especialmente en entornos cambiantes. Permite mejorar la eficiencia, generar ventajas competitivas, resolver problemas y fomentar la creatividad institucional, siendo clave para el progreso organizacional (Findikoğlu y İlhan, 2016).

En educación, Halasz (2021) define la innovación pedagógica como el acto de crear y difundir nuevas herramientas educativas, que rompen con las prácticas tradicionales. Walder (2017) enfatiza en que la innovación educativa debe involucrar una práctica docente que pretenda mejorar el aprendizaje, lo que para Krstikj et al. (2022), implica el logro de los resultados de aprendizaje e impactar positivamente el desempeño estudiantil.

En este sentido, la innovación docente en educación superior busca transformar los procesos educativos tradicionales para adaptarlos a las necesidades y demandas de la sociedad, fomentando el desarrollo de habilidades y competencias relevantes para el mundo laboral (Sola Fernández, 2016). Consecuentemente, se ha enfatizado en la necesidad de una cultura basada en la innovación en la práctica docente en educación superior (Macanchí Pico et al., 2020).

Actualmente, la sociedad demanda que las instituciones académicas sean más flexibles, y comprometidas con el desarrollo sostenible de sus comunidades. Por ejemplo, que sus programas educativos y áreas de investigación sean pertinentes, alineados con las necesidades del mundo laboral y la sociedad, fomentando la innovación, el espíritu emprendedor y el pensamiento crítico, garantizando el acceso e inclusión a la educación superior a todos los sectores de la sociedad, promoviendo la transferencia de conocimientos y la colaboración con y entre distintos actores sociales (den Brok, 2018).

Los egresados deben ser capaces de operar en entornos complejos, inciertos y dinámicos (Villaruel y Bruna, 2014), relevando importancia al desarrollo de habilidades transversales. Esto tiene implicancias sobre qué y cómo enseñar. Los métodos innovadores de enseñanza-aprendizaje son cruciales para promover un espíritu de aprendizaje continuo (Lotz-Sisitka et al., 2015). Consecuentemente, López (2022) plantea que las universidades tienen la responsabilidad de comprometer y motivar al estudiantado utilizando métodos de aprendizaje centrados en el estudiantado y en el aprendizaje colaborativo.

La educación superior enfrenta el desafío de superar modelos transmisivos, promoviendo la participación activa, la interacción y co-construcción de conocimientos, habilidades y valores (Burganova et al., 2018). Cobran relevancia las experiencias de aprendizaje experiencial y auténtico (Amante y Fernandes, 2023), considerando tanto procesos de enseñanza presencial, como e-learning y b-learning (Chandra et al., 2021). Para ello, las instituciones capacitan al profesorado y ofrecen instrumentos para promover la innovación educativa, tales como fondos de apoyo e incentivos (Kezar y Eckel, 2002).

Existen distintas clasificaciones de innovación educativa: curricular, didáctica, evaluativa o tecnológica. También distintos niveles de intervención: estudiantes, docentes, programa u organización. Su implementación requiere planificar y evaluar su coherencia con el contexto, los resultados de aprendizaje, competencias del profesorado y estudiantado, y los recursos disponibles, dado que la planificación es tan relevante como su ejecución y evaluación (Marchant et al., 2018).

La evaluación de una innovación es fundamental para comprender su efectividad, identificar áreas de mejora y tomar decisiones informadas (Lin et al., 2022). Más allá de medir éxito o fracaso, la evaluación busca comprender y mejorar los procesos de enseñanza-aprendizaje mediante indicadores que permitan valorar el logro de los resultados deseados (Cook et al., 2022).

La evaluación debe considerar los objetivos y la novedad de la innovación, su comparación con enfoques tradicionales y su impacto real en el aprendizaje, utilizando una variedad de métodos (Halász, 2021; Mazandarani y Troudi, 2022).

La información sobre la innovación puede obtenerse mediante métodos mixtos (encuestas, observaciones, entrevistas, análisis de desempeño), incorporando la percepción de docentes y estudiantes sobre su valor y efectos en el aprendizaje (Kauffeld et al., 2019). Esto acerca la innovación a la investigación educativa, que ha demostrado mejorar la autoeficacia y proactividad docente, impactando, por ejemplo, la enseñanza del emprendimiento (Galindo-Domínguez et al., 2022).

Se ha encontrado que la calidad percibida de una innovación se relaciona positivamente con la satisfacción estudiantil y con el enfoque pedagógico del docente y la asignatura (Calma y Dickson-Deane, 2020). Lin et al. (2022) reportaron que la innovación docente se asoció significativamente con la evaluación de los estudiantes, explicando casi el 40% de sus resultados, destacando el papel central del estudiante en su evaluación.

Respecto a los desafíos en la evaluación del impacto, es necesario avanzar hacia sistemas más objetivos y fiables que integren la planificación, implementación y retroalimentación de la innovación, complementados con metodologías cualitativas que caractericen la innovación y profundicen en su significado desde la experiencia del profesorado y estudiantado (Xin et al., 2022).

En este contexto, se ha planteado que la innovación educativa es un campo emergente que presenta el desafío de desarrollar estrategias de evaluación efectivas estandarizadas (Chandra et al., 2021). Sin embargo, las innovaciones educativas acostumbra a carecer de una validación empírica (Pedró, 2015), debido en parte a la limitada formación investigativa del profesorado, tradicionalmente seleccionado por su experticia disciplinar. Una forma de abordar esta brecha es mediante protocolos e instrumentos confiables.

La literatura reporta la creación y validación psicométrica de escalas relacionadas con la innovación pedagógica sobre actitudes (Traver-Martí y Fernández-Barrueco 2016), prácticas (Santos Rego et al., 2017), uso de tecnología (López-Pérez et al., 2023), condiciones institucionales (Cifuentes y Herrera Helásquez, 2019) y competencias docentes para la innovación (Dueñas Zorrilla et al., 2024), las que se centran en el profesorado. De estas, incluso la escala de Traver-Martí y Fernández-Barrueco (2016), que contempla una versión para estudiantes, incluye dimensiones y reactivos orientados al rol del docente y factores externos al estudiante. Respecto a la evaluación desde la perspectiva estudiantil, existe escasa validación de instrumentos que midan la percepción sobre innovaciones diversas. En este contexto, Pozuelos Estrada et al. (2021) desarrollaron y validaron psicométricamente una escala para evaluar prácticas innovadoras desde la perspectiva estudiantil, pero específicamente orientada al desarrollo de proyectos e investigación, con reactivos enfocados a recoger la preferencia personal en los aspectos de diseño (Estrategia, Material, Apoyo y Evaluación). En consecuencia, se identifica la necesidad de contar con instrumentos validados que recojan la experiencia del estudiantado en la evaluación de innovaciones educativas en contextos diversos, foco de este estudio.

## 2. Metodología

### 2.1. Objetivos

Este trabajo tiene por objetivo crear y validar una escala para evaluar innovaciones educativas de amplio uso, aplicable a la diversidad de diseños innovadores, que facilite la toma de decisiones basada en evidencia para mejorar el proceso enseñanza-aprendizaje.

### 2.2. Participantes

Se utilizó un muestreo no probabilístico por conveniencia, seleccionando aquellos estudiantes que aceptaron ser incluidos, de acuerdo con la accesibilidad y proximidad de los sujetos con el investigador (Otzen y Manterola, 2017). Como resultado, participaron 450 estudiantes de pregrado pertenecientes a 22 carreras de 8 facultades de una universidad privada chilena, que utilizaron la escala para evaluar una intervención pedagógica en alguna de sus asignaturas cursadas.

La muestra estuvo compuesta por un 69,3% de mujeres (N = 312) y un 30,7% de hombres (N = 138), con edades entre los 17 y 36 años (M = 20,19; DE = 2,68). La Tabla 1 detalla la cantidad de intervenciones y estudiantes que evaluaron dichas intervenciones, según la clasificación de áreas del conocimiento de la OECD (Organization for Economic Cooperation and Development, 2015).

**Tabla 1**

*Distribución de la muestra según área de conocimiento*

Área del conocimiento	Nº de intervenciones	Nº de estudiantes
Ciencias naturales	8	77
Ingeniería y tecnología	4	112
Ciencias médicas y de la salud	9	140
Ciencias sociales	9	95
Humanidades	1	26

### 2.3. Instrumento

La construcción del instrumento se llevó a cabo a través de varias etapas consecutivas. Se diseñó originalmente de 40 ítems. El análisis de panel de expertos (Davis, 1992), resultó en la eliminación de 15 ítems que no cumplieron el umbral de RVC (Razón de Validez de Contenido) de 0,6, criterio adecuado para paneles de 10 expertos (Lawshe, 1975). Como resultado, la versión piloto se compuso de 25 ítems.

#### 2.3.1. Versión preliminar del Cuestionario de Percepción Sobre la Implementación de una Intervención Pedagógica en el Aula

El diseño del instrumento se realizó mediante trabajo colaborativo por distintos profesionales expertos en elaboración de instrumentos, innovación educativa y evaluación. La propuesta original consistió en un instrumento de 40 ítems, con respuesta tipo Likert de cinco puntos, que abordó cinco dimensiones descritas en la Tabla 2.

**Tabla 2**

*Descripción de las dimensiones*

Dimensión	Descripción
1. Relación entre la metodología de enseñanza empleada y resultados de aprendizaje de la asignatura.	Grado en que el estudiante considera que la metodología empleada ha influido positivamente en la calidad de los aprendizajes consolidados.
2. Relación entre la metodología de enseñanza y el desarrollo de competencias genéricas.	Grado en que el estudiante percibe que la metodología empleada favorece el desarrollo de habilidades y actitudes relacionadas con el desarrollo de competencias genéricas.
3. Contribución de la metodología en la motivación del estudiante por el aprendizaje.	Grado en que el estudiante percibe que la metodología impactó positivamente en la motivación y/o compromiso que siente hacia los saberes disciplinares y su carrera.
4. Percepción del diseño y la planificación de la metodología.	Grado en que el estudiante percibe que la metodología fue adecuadamente diseñada y planificada en cuanto a pertinencia y calidad, y la preferencia de la metodología en comparación a otras metodologías tradicionales.
5. Percepción de la ejecución de la metodología por parte del docente.	Grado en que el estudiante percibe positivamente el rol y competencia del profesor en la implementación de la metodología de enseñanza.

#### 2.3.2. Evaluación de ítems por panel de expertos

La propuesta original de la escala fue sometida a juicio experto (Davis, 1992). Se contó con la participación de 10 expertos, tres expertos en evaluación e instrumentos, y siete con vasta experiencia en intervenciones pedagógicas. Los expertos fueron contactados mediante correo electrónico, adjuntando instrucciones y la pauta de valoración, recibiendo sus comentarios por el mismo medio. Se empleó una pauta que describía las dimensiones abordadas y sus definiciones teóricas, y los 40 ítems de la versión preliminar. La valoración de los ítems se realizó valorando cada ítem como “esencial”, «útil pero no esencial», o «no necesario» en el marco de la dimensión correspondiente. Adicionalmente, los expertos podían realizar comentarios cualitativos sobre la redacción o composición de los reactivos.

### 2.3.3. Versión piloto del cuestionario

La retroalimentación recibida del juicio de experto resultó en la eliminación de 15 ítems y la revisión de la redacción de los reactivos. Como resultado, el instrumento piloto quedó conformado por 25 ítems (Bruna et al., 2020) distribuidos en las mismas cinco dimensiones de la versión preliminar (tabla 3).

**Tabla 3**

*Escala resultante del juicio de expertos*

Dimensión	Ítems
1	1. La metodología favoreció la comprensión de los contenidos de la asignatura. 2. La metodología me facilitó la integración de contenidos de la asignatura. 3. La metodología me permitió aclarar dudas sobre los contenidos de la asignatura. 4. La metodología me permitió aplicar los contenidos en el aula. 5. La metodología me permitió entender la utilidad de los contenidos en mi futuro quehacer profesional. 6. La metodología me entregó retroalimentación sobre lo aprendido.
2	7. La metodología promovió el trabajo en equipo. 8. La metodología favoreció la interacción con mis pares. 9. La metodología favoreció mi interacción con el/la profesor/a. 10. La metodología promovió mi análisis sobre el contenido. 11. La metodología promovió mi reflexión sobre mi aprendizaje. 12. La metodología facilitó mi comunicación oral. 13. La metodología favoreció mi comunicación escrita.
3	14. La metodología me motivó a estudiar. 15. La metodología contribuyó a que valorara mi rol como futuro profesional. 16. La metodología promovió que me sintiera responsable de mi aprendizaje.
4	17. La metodología me parece adecuada para lograr los resultados de aprendizaje de la asignatura. 18. El material y/o recursos utilizados me parecieron apropiados y de calidad para el logro de los aprendizajes. 19. El tiempo destinado al desarrollo de la/s actividad/es fue suficiente. 20. Prefiero una clase/evaluación tradicional que esta estrategia. 21. Me gustaría que esta metodología se utilizara en otras asignaturas.
5	22. El objetivo de la metodología fue explicado claramente. 23. El profesor demostró dominio de la metodología implementada. 24. Durante la implementación de la metodología, las dudas fueron resueltas oportunamente. 25. La guía del profesor en la implementación de la metodología contribuyó a mi aprendizaje.

## 2.4. Procedimiento

Se contactó a los estudiantes por correo institucional coordinando previamente con los docentes responsables de las asignaturas con innovaciones. Las innovaciones involucraban nuevas metodologías de enseñanza y evaluación de aprendizajes, tales como: actividades activas durante laboratorios vinculando imágenes con el contenido, uso de infografías de contenido para evaluar aprendizajes, ejercicios o casos auténticos para conectar con los contenidos y resultados de aprendizajes, uso de recursos y aprendizajes en línea que acompañaban las sesiones presenciales, entrega de retroalimentación, actividades de aplicación o de toma de decisiones grupales, entre otros. La escala se aplicó en línea, considerando un periodo de dos semanas para contestar. Todas las respuestas se consideraron obligatorias, por lo que no hay datos faltantes. Se comprometió la entrega de los resultados obtenidos sin la identificación de los estudiantes para retroalimentar al docente responsable. El proceso fue realizado acorde con los principios éticos del Comité de Ética de Investigación de la universidad participante.

## 2.5. Análisis de datos

Se realizó un análisis descriptivo como paso previo al análisis factorial confirmatorio. Dado que la estructura del instrumento tenía una base teórica sólida, pero podía requerir ajustes específicos en su estructura factorial, se decidió dividir la muestra aleatoriamente en dos submuestras y aplicar Modelos de Ecuaciones Estructurales (SEM) en ambas. En la primera submuestra (60% de los casos), se evaluó el ajuste de tres modelos, permitiendo la búsqueda de especificación para mejorar la estructura factorial a través de un análisis factorial semi-confirmatorio. En la segunda submuestra (40% de los casos), se llevó a cabo un análisis factorial confirmatorio estricto, considerando únicamente los modelos que mostraron un buen ajuste en la primera submuestra.

Se estudiaron cuatro modelos de estructura factorial, compatibles con la estructura teórica del instrumento: 1) estructura unidimensional, 2) modelo de cinco dimensiones relacionadas, 3) modelo jerárquico de cinco dimensiones, 4) modelo bifactorial (SI-1), con el primer ítem de la primera dimensión considerada como referencia. Se utilizó el modelo bifactorial (SI-1) ya que se considera que los factores no son intercambiables, sino que reflejan factores con significado único, lo que evita problemas de estimación posteriores (Eid et al., 2017). En función del análisis del modelo SI-1, se consideró apropiado reformularlo como S-1, estableciendo una de las dimensiones teóricas como referencia, ya que las cargas factoriales de esta dimensión no resultaron significativas. Se utilizó estimación WLSMV, adecuada para escalas conformadas por ítems a nivel ordinal de cinco puntos o menos. Se evaluó cada modelo usando Chi-cuadrado normalizado ( $\chi^2/gl$ ), CFI, TLI, RMSEA y SRMR. Los criterios iniciales usados para evaluar el modelo fueron: (a)  $\chi^2/gl < 3$ , (b) CFI y TLI mayores a 0,9 corresponden a un ajuste aceptable y sobre 0,95 a un buen ajuste, (c) RMSEA con valores menores a 0,08 indican un ajuste aceptable y menores a 0,06 un buen ajuste, (d) SRMR con valores menores a 0,08 indica un buen ajuste de los datos al modelo (Hu y Bentler, 1999). Considerando que los criterios de ajuste deben adaptarse al tipo de ítems, estructura propuesta y tipo de estimación (Xia y Yang, 2019), se utilizó la metodología Dynamic Fit Index (DFI), que genera indicadores análogos a los de Hu y Bentler, simulando en función del tamaño de muestra y modelo analizado (McNeish y Wolf, 2023). Usando la simulación para variables categóricas usando estimación WLSMV, los valores del percentil 99% para el modelos unifactorial y el modelo de 5 factores fueron idénticos, de SRMR = 0,031, RMSEA = 0,024 y CFI = 0,998. Se debe considerar que, en la simulación de 5 factores relacionados, se encontró que el tamaño de muestra no permite distinguir de manera adecuada entre el modelo correcto y modelos con problemas de especificación de cargas cruzadas.

Se consideró como solución factorial aceptable aquella que presentase buenos índices de ajuste, así como cargas factoriales de al menos 0,4 y significativas en todos los ítems. Como los modelos alternativos para el análisis confirmatorio no eran anidados y no es posible usar AIC y BIC, ya que el estimador WLSMV no utiliza la verosimilitud como criterio, se estimaron los modelos con MLR para poder obtener estos indicadores.

La confiabilidad de las escalas fue analizada tanto para el caso individual, como para la evaluación de la asignatura. El análisis de la confiabilidad de la respuesta a nivel individual se realizó considerando tanto el coeficiente alfa de Cronbach, como omega de McDonald. Para el análisis de la confiabilidad de la medida a nivel de asignatura, se calculó el ICC y se realizó un estudio de Generalizability. En este, se realizó primero un estudio G para analizar las fuentes de varianza considerando modelos lineales mixtos jerárquicos, considerando como población a ser evaluada las asignaturas, y como facetas los ítems y los estudiantes como efectos aleatorios. Luego, se realizó un estudio D para determinar en qué medida se modifica la confiabilidad de la medida al modificar el número de estudiantes por asignatura, calculando el coeficiente de generabilidad  $\rho$  (generalizability coefficient).

Todos los análisis se realizaron usando el programa R, versión 4.5. Se usaron los paquetes psych, para hacer descripción de manera general, lavaan y semTools para realizar SEM, y dynamic para calcular los valores de corte para CFI, SRMR y RMSEA.

### 3. Resultados

Los estadísticos de los ítems de la escala se resumen en la Tabla 4. La mayoría de los ítems presentan valores de asimetría o curtosis que exceden el rango [-1, +1]. Este patrón puede atribuirse a las altas medias observadas en los reactivos del cuestionario, acompañadas de asimetrías negativas y curtosis positivas elevadas, sugiriendo un alto nivel de acuerdo con los ítems evaluados.

**Tabla 4**  
Descripción de los ítems de la escala

Nº ítems	Dimensión	Media	D.E.	Asim.	Curtosis
1	1	4,22	1,09	-1,67	2,22
2		4,20	1,07	-1,59	2,04
3		4,16	1,05	-1,42	1,57
4		4,06	1,09	-1,25	1,02
5		4,27	0,99	-1,67	2,70
6		4,18	-0,97	-1,50	1,85
7	2	4,03	1,06	-0,97	0,25
8		3,96	1,07	-0,95	0,30
9		4,03	1,14	-1,16	0,57
10		4,27	0,98	-1,70	2,82
11		4,24	1,03	-1,54	1,99
12		3,46	1,25	-0,39	-0,85
13		4,08	1,00	-1,11	0,96

Nº ítems	Dimensión	Media	D.E.	Asim.	Curtosis
14	3	3,88	1,17	-1,10	0,42
15		4,18	1,02	-1,44	1,80
16		4,38	0,88	-1,92	4,28
17	4	4,05	1,15	-1,23	0,72
18		4,20	1,04	-1,53	1,95
19		4,13	1,17	-1,38	1,04
20 <sup>a</sup>		3,19	1,33	-0,31	-1,02
21		3,94	1,18	-1,12	0,44
22	5	4,24	1,05	-1,64	2,32
23		4,48	0,86	-2,19	5,31
24		4,31	0,96	-1,56	2,21
25		4,32	1,03	-1,73	2,52

Nota. N = 450. Todos los ítems fueron contestados en escala de 1 al 5.<sup>a</sup> Ítem codificado de forma inversa. Eliminado de la solución factorial final.

### 3.1. Análisis Factorial Confirmatorio

Se probaron cuatro modelos con la primera submuestra. Los tres primeros modelos surgieron directamente de la estructura teórica que presenta el instrumento: modelo unidimensional (M0), modelo de cinco factores relacionados (M1), modelo jerárquico de cinco factores (M2) y modelo bifactorial (M3). A partir de la búsqueda de especificación del modelo con mejor ajuste, se probó un modelo más, el modelo bifactorial corregido (M4). En las primeras iteraciones de todos los modelos, se encontró que el ítem 20, el único ítem invertido de todo el instrumento no contaba con la carga factorial suficiente ( $>0,4$ ), por lo que fue eliminado y se ajustaron todos los modelos sin este ítem. Esto implica que todos ellos tienen un carácter exploratorio, siendo pertinente el análisis factorial confirmatorio en una segunda muestra.

Como se observa en la Tabla 5, el modelo unidimensional (M0) presenta indicadores no favorables, al tener  $\chi^2/g.l > 3$ , SRMR  $> 0,08$  y RMSEA  $> 0,08$ , por lo que se descarta completamente. Tanto los modelos de cinco factores relacionados (M1), como el jerárquico (M2) presentan mejores indicadores de ajuste respecto al unidimensional al mostrar CFI y TLI cercanos a 0,95, y SRMR  $< 0,08$ . No obstante, ambos presentan  $\chi^2/g.l$  mayor a 3 y RMSEA  $> 0,08$ , indicando que ambos modelos presentan un nivel de error relevante con criterios de corte tradicionales. Esto se confirma al usar como criterio los valores de la simulación DFI, ya que ambos modelos presentan indicadores que no cumplen con los criterios mínimos, de SRMR  $< 0,031$ , RMSEA  $< 0,024$  y CFI  $> 0,998$ . El modelo bifactorial (M3), por su parte, presenta buenos indicadores de CFI y TLI, cercanos a 0,99, lo que satisface el criterio de la simulación DFI. A pesar de que el SRMR = 0,038 está sobre el valor de corte 0,031 y RMSEA = 0,066 es superior al valor de corte 0,024 de la simulación DFI, se encuentran relativamente cercanos y cumplen con los criterios tradicionales de ajuste. Por tanto, se analizó la estructura de cargas factoriales y los índices de modificación para evaluar si es posible realizar una búsqueda de especificación del modelo bifactorial (M3).

Al analizar las cargas factoriales del factor general del modelo bifactorial, se puede observar que están superan el 0,4, excepto los ítems 7 y 8, pero en ambos casos las cargas factoriales son 0,39, lo que reafirma el carácter fundamentalmente unidimensional de la escala. Los índices de modificación más relevantes son una relación entre los residuos de los ítems 1 y 2 (IM = 22,84) y la correlación entre las dimensiones 4 y 5 (IM = 14,04), los cuales son menores al 10% del chi-cuadrado total, por lo que no se consideran suficientemente grandes para establecer. Sin embargo, al analizar la dimensión 3 'Contribución de la metodología en la motivación por el aprendizaje', se puede observar que ninguna de las cargas factoriales son significativas. Por tanto, se consideró ajustar un modelo bifactorial S-1, donde la dimensión 3 se considera de referencia, eliminándose las cargas factoriales grupales de esta dimensión. Este modelo presenta buenos indicadores de ajuste,  $\chi^2 (171) = 233,96$ , CFI = 0,995, SRMR = 0,034 y RMSEA = 0,045, ya que usando como referencia la simulación DFI, se cumple el criterio para CFI, y los valores de SRMR y RMSEA son cercanos a los valores de referencia. Al considerar estos indicadores y que el índice de modificación mayor, correspondiente a la relación entre residuos de los ítems 10 y 11 es pequeño (IM = 9,28), se considera al modelo 4 apropiado para ser analizado.

Tabla 5

Modelos del análisis factorial confirmatorio, submuestra 1

Modelo	X <sup>2</sup>	X <sup>2</sup> /gl	CFI	TLI	SRMR	RMSEA
<b>M0: Unidimensional</b>	X <sup>2</sup> (252) = 1514,57 p < 0,001	6,01	0,944	0,938	0,08	0,136 [0,130, 0,143], p < 0,001
<b>M1: 5 Factores relacionados</b>	X <sup>2</sup> (242) = 1399,72 p < 0,001	5,784	0,948	0,941	0,074	0,133 [0,127, 0,140], p < 0,001
<b>M2: Jerárquico</b>	X <sup>2</sup> (247) = 1318,41 p < 0,001	5,338	0,952	0,946	0,075	0,127 [0,120, 0,134], p < 0,001
<b>M3: Bifactorial SI-1</b>	X <sup>2</sup> (229) = 499,56 p < 0,001	2,181	0,988	0,985	0,038	0,066 [0,058, 0,074], p = 0,000
<b>M4: Bifactorial S-1</b>	X <sup>2</sup> (231) = 475,93 p < 0,001	2,06	0,989	0,987	0,038	0,063 [0,055, 0,071], p = 0,005

Notas: El modelo bifactorial SI-1 utiliza el primer ítem como referencia. El modelo bifactorial S-1 utiliza la dimensión motivacional como referencia.

Al probar el ajuste de los modelos M3 y M4 en la segunda submuestra usando estimación WLSMV, se observa en la Tabla 6 que los resultados son similares obtenidos en la muestra 1 con respecto a los indicadores de ajuste, ya que se cumple el criterio basado en DFI para CFI, pero no para SRMR y RMSEA. Hay una pequeña diferencia a favor del modelo M4 respecto al M3 en todos los indicadores,  $\Delta CFI = 0,002$ ,  $\Delta SRMR = 0,001$  y  $\Delta RMSEA = 0,005$ . Al analizar la estimación MLR, la diferencia en indicadores de ajuste a favor del modelo M4 es mayor en  $\Delta CFI = 0,007$  y  $\Delta SRMR = 0,009$ , aunque menor en  $\Delta RMSEA = 0,004$ . Los indicadores AIC y BIC son menores en M4. Considerando que el modelo M4 es más parsimonioso, presenta mejores indicadores de ajuste y presenta un claro sentido teórico, destacando el papel de la intervención para el desarrollo de la motivación de la asignatura y la carrera, se selecciona este modelo como adecuado.

Tabla 6

Modelos del análisis factorial confirmatorio, submuestra 2

Modelo	X <sup>2</sup>	X <sup>2</sup> /gl	CFI	TLI	SRMR	RMSEA	AIC	BIC
<b>WLSMV</b>								
<b>M3: Bifactorial SI-1</b>	X <sup>2</sup> (229) = 358,97 p < 0,001	1,568	0,992	0,991	0,042	0,056 [0,045, 0,067], p = 0,174	—	—
<b>M4: Bifactorial S-1</b>	X <sup>2</sup> (231) = 336,77 p < 0,001	1,458	0,994	0,993	0,041	0,051 [0,038, 0,062], p = 0,457	—	—
<b>MLR</b>								
<b>M3: Bifactorial SI-1</b>	X <sup>2</sup> (229) = 405,9 p < 0,001	1,771	0,939	0,926	0,041	0,065 [0,056, 0,074], p = 0,003	8860,46	9087,16
<b>M4: Bifactorial S-1</b>	X <sup>2</sup> (231) = 387,43 p < 0,001	1,667	0,946	0,935	0,040	0,061 [0,052, 0,070], p = 0,023	8834,52	9054,83

Notas: El modelo bifactorial SI-1 utiliza el primer ítem como referencia. El modelo bifactorial S-1 utiliza la dimensión motivacional como referencia.

La Figura 1 muestra el modelo bifactorial S-1, que considera 1 factor general y 4 factores específicos. Todas las cargas factoriales del factor general G son superiores a 0,4, excepto I7, pero su intervalo de confianza al 95% [0,29, 0,46] incluye el 0,4, por lo que todos los ítems cargan de manera suficiente en G. Por su parte, todas las cargas factoriales de los factores específicos son significativas, excepto por I11.

Para evaluar el carácter unidimensional o bifactorial de la solución, se debe considerar que, en primer lugar, la varianza explicada por el factor G es 86,6%, por lo que la varianza de los factores específicos es 13,4 %. El indicador PUC, que mide el grado en que las correlaciones no están afectas a multidimensionalidad, es de 82,6%. La carga promedio de las cargas factoriales de los factores específicos es 0,28, lo que está bajo del umbral de 0,3 que se utiliza para declarar la necesidad de un modelo bifactorial (Sánchez-Villena

et al., 2021). Todo esto indica que la escala se debe tratar como fundamentalmente unidimensional, siendo necesario considerar los factores específicos para explicar la varianza residual que no explica G.

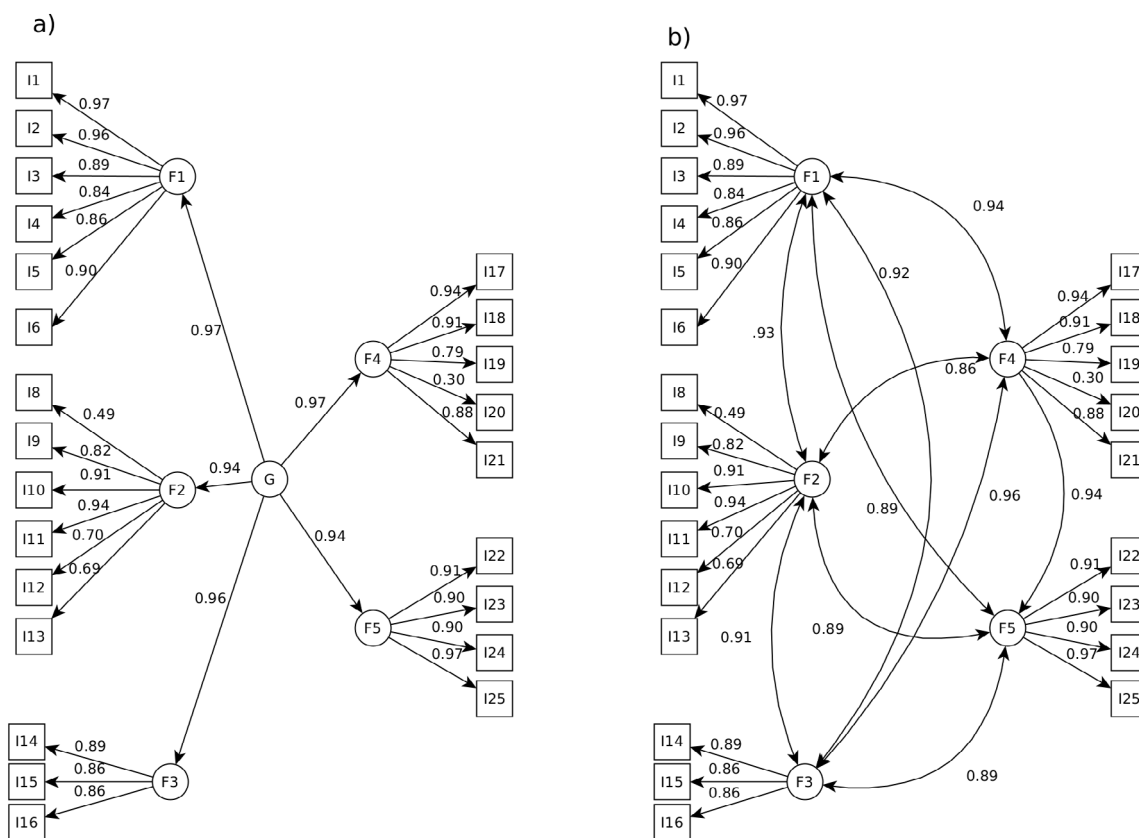


Figura 1. Modelo de análisis factorial confirmatorio en la muestra total.

### 3.2. Análisis descriptivo y de confiabilidad de las escalas

La Tabla 7 muestra los descriptivos y los indicadores de confiabilidad como consistencia interna para las cinco escalas. La escala con los menores puntajes fue Factor 2 y Factor 4 para ambas muestras, mientras que los puntajes más altos se presentaron en el Factor 5. La confiabilidad de la escala general es muy buena, con valores de alfa = 0,97 y omega total = 0,98. El omega jerárquico, que mide la confiabilidad del factor general eliminando el efecto de los factores específicos, es también alto, omega<sub>j</sub> = 0,95.

Tabla 7

Descriptivos e indicadores de confiabilidad para los cinco factores de la escala de innovación docente

Factor	Media	D.E.	Asim.	Kr	alfa	omega <sub>T</sub>	omega <sub>j</sub>	ICC
1. RA	4,181	0,9418	-1,678	2,662	0,95	0,94	0,041	0,39
2. CG	4,007	0,804	-1,097	1,65	0,87	0,85	0,214	0,28
3. MA	4,144	0,9078	-1,524	2,323	0,86	0,84	—	0,33
4. DP	4,079	0,9984	-1,371	1,349	0,90	0,82	0,042	0,47
5. ED	4,338	0,8804	-1,850	3,389	0,92	0,91	0,087	0,47
<b>Total</b>	4,135	0,8210	-1,544	2,417	0,97	0,98	0,946	0,42

Notas: RA: Resultados de aprendizaje. CG: Competencias genéricas. MA: Motivación por el aprendizaje (dimensión de referencia). DP: Diseño y planificación. ED: Ejecución de metodología por parte del docente. omega<sub>T</sub>: Omega total. Omega<sub>j</sub>: Omega jerárquico. No se calcula omega jerárquico para la dimensión 3, ya que es la dimensión de referencia para el modelo bifactorial S-1.

Al analizar correlación intraclase por asignatura, correspondiente al grado en que las respuestas se relacionan entre sí dentro de cada asignatura, se aprecia que oscilan entre 0,28 en CG y 0,47 para DP y ED, lo que corresponde a una asociación fuerte. Al realizar el estudio G, que descompone las fuentes de variabilidad de la respuesta de cada ítem para cada individuo, se puede encontrar la mayor fuente de variación corresponde a la variabilidad entre los estudiantes dentro de cada asignatura (0,390; 34,4% de la varianza), siguiéndole el error de medida no explicado en cada ítem (0,386; 34,1% de la varianza), y las diferencias entre asignaturas (0,23; 20,2% de la varianza). Las fuentes más pequeñas de varianza corresponden al comportamiento diferencial de los ítems por asignatura (0,089; 7,8% de la varianza), así como la varianza de los ítems (0,040; 3,5% de la varianza). La baja variabilidad asociada a la interacción ítem X carrera indica que el instrumento se comporta de forma estable entre carreras. Sin embargo, el alto nivel de varianza asociado a los estudiantes dentro de cada carrera, así como el error de medida no explicada, indica que la medición confiable de los valores de una asignatura requiere una cantidad mínima de estudiantes por asignatura. El estudio D indica que el coeficiente de generabilidad para un estudiante es  $\rho = 0,36$ , para  $N = 4$  es  $\rho = 0,70$ , para  $N = 7$  es  $\rho = 0,80$  y para  $N = 16$  es  $\rho = 0,90$ . Esto indica que para que una medición permita realizar decisiones sobre asignaturas específicas, se debe contar con la medición de al menos 16 estudiantes.

### 3.3. Escala final

La escala se consolidó como un instrumento de 24 ítems organizados en los cinco factores originales, con el ítem 20 eliminado («Prefiero una clase/evaluación tradicional que esta estrategia»). Cabe señalar, que los factores se renombraron en conceptos que representan la definición original y las definiciones incluyen cambios menores para precisar cada dimensión.

La escala final, renombrada Escala de Evaluación de Innovaciones Educativas (EVINED), y renumerados los ítems, se presenta en la Tabla 8.

**Tabla 8**

*Escala de Evaluación de Innovaciones Educativas (EVINED)*

Dimensión	Ítems
<p><b>1. Promoción de aprendizaje disciplinar:</b> Grado en que el estudiante considera que la metodología empleada (innovación realizada) ha influido positivamente en la calidad de los aprendizajes adquiridos.</p>	<p>1. La metodología favoreció la comprensión de los contenidos de la asignatura. 2. La metodología me facilitó la integración de contenidos de la asignatura. 3. La metodología me permitió aclarar dudas sobre los contenidos de la asignatura. 4. La metodología me permitió aplicar los contenidos en el aula. 5. La metodología me permitió entender la utilidad de los contenidos en mi futuro quehacer profesional. 6. La metodología me entregó retroalimentación sobre lo aprendido.</p>
<p><b>2. Promoción de habilidades transversales:</b> Grado en que el estudiante percibe que la metodología empleada favorece el desarrollo de habilidades, destrezas y actitudes que se relacionan con competencias genéricas.</p>	<p>7. La metodología promovió el trabajo en equipo. 8. La metodología favoreció la interacción con mis pares. 9. La metodología favoreció mi interacción con el/la profesor/a. 10. La metodología promovió mi análisis sobre el contenido. 11. La metodología promovió mi reflexión sobre mi aprendizaje. 12. La metodología facilitó mi comunicación oral. 13. La metodología favoreció mi comunicación escrita.</p>
<p><b>3. Motivación:</b> Grado en que el estudiante percibe que la metodología impactó positivamente en la motivación y/o compromiso que siente hacia los saberes disciplinares y su carrera.</p>	<p>14. La metodología me motivó a estudiar. 15. La metodología contribuyó a que valorara mi rol como futuro profesional. 16. La metodología promovió que me sintiera responsable de mi aprendizaje.</p>
<p><b>4. Diseño:</b> Grado en que el estudiante percibe que la metodología fue adecuadamente planificada en términos de tributar a la promoción del logro de aprendizajes esperados del programa y otros aspectos.</p>	<p>17. La metodología me parece adecuada para lograr los resultados de aprendizaje de la asignatura. 18. El material y/o recursos utilizados me parecieron apropiados y de calidad para el logro de los aprendizajes. 19. El tiempo destinado al desarrollo de la/s actividad/es fue suficiente. 20. Me gustaría que esta metodología se utilizara en otras asignaturas.</p>
<p><b>5. Implementación:</b> Grado en que el estudiante percibe positivamente el rol y competencia del profesor en la implementación de la innovación.</p>	<p>21. El objetivo de la metodología fue explicado claramente. 22. El profesor demostró dominio de la metodología implementada. 23. Durante la implementación de la metodología, las dudas fueron resueltas oportunamente. 24. La guía del profesor en la implementación de la metodología contribuyó a mi aprendizaje.</p>

En cuanto a los resultados de la aplicación de la escala a la muestra se obtuvo un alto grado de acuerdo parcial y total para todas las dimensiones (promedio de los ítems 74 a 87%). La dimensión con menor acuerdo corresponde a la Promoción de habilidades transversales (promedio de los ítems 74%), en la cual destaca el bajo acuerdo con respecto a facilitar la comunicación oral (52%). La dimensión con mayor acuerdo corresponde a la Implementación (promedio de los ítems 87%), en la cual destaca el alto acuerdo con respecto al dominio metodológico del profesor (92%). Otros hallazgos para destacar es el alto acuerdo sobre que la metodología promueve que se sientan responsables de su aprendizaje, de la dimensión Motivación (90%), y la alta y consistente percepción de acuerdo de todos los ítems de la dimensión Promoción del aprendizaje disciplinar (79-84%).

#### 4. Discusión y conclusiones

Las necesidades del campo laboral requieren de la integración de competencias disciplinares y transversales (Villaruel y Bruna, 2014). Por otra parte, el perfil actual del estudiantado, más diverso, con brechas en conocimientos y habilidades producto de acceso a la educación superior aumentado por políticas nacionales (Henríquez y Vargas, 2022) y deprimido debido a la pandemia (Betthäuser et al., 2023), interpela a implementar innovaciones en el proceso enseñanza-aprendizaje.

Las instituciones de educación superior apoyan este proceso a través de capacitaciones, fondos concursables e incentivos (Kezar y Eckel, 2002). Aunque el profesorado acceda a dichos apoyos e implemente innovaciones en el aula, en su mayoría carecen de los conocimientos necesarios para crear instrumentos confiables para evaluar dichas intervenciones y basar las decisiones pedagógicas en evidencias confiables. En este contexto, este trabajo buscó generar y poner a disposición un instrumento validado psicométricamente para evaluar las innovaciones implementadas en el aula, cualquiera sea su naturaleza.

El trabajo comenzó desde una escala creada y validada por juicio de experto de 25 ítems organizados en cinco dimensiones presentada como versión piloto (Bruna et al., 2020), la cual se utilizó en asignaturas en las que se implementaron innovaciones en diversas áreas del conocimiento, de acuerdo con la Tabla 1, dando cuenta de la generalidad de su potencial uso.

En la primera parte del estudio se comprobó que los modelos teóricos iniciales no se ajustaban bien a los datos lo que llevó a revisar la estructura de la escala y eliminar el ítem 20: «Prefiero una clase evaluación tradicional que esta estrategia». La interpretación de este ítem puede ser complejo en el sentido de diferenciar lo tradicional respecto de lo innovador. Con esta corrección se analizaron los modelos obteniendo indicadores adecuados con el modelo bifactorial, el que aún presentaba necesidades de ajuste, por lo que se realizó una búsqueda de especificación. Las cargas factoriales sugirieron considerar la dimensión 3 «Contribución de la metodología en la motivación por el aprendizaje» como referencia, modelo que presentó mejores indicadores y resultó más parsimonioso, seleccionándolo como adecuado. Este resultado sugiere a futuro fortalecer la dimensión 3 con más ítems relativos al tema para una medición más precisa.

Cabe señalar que el factor general explica la mayor parte de la varianza, combinando un buen ajuste estadístico con una consistencia teórica, por lo que la escala se debe tratar fundamentalmente como unidimensional, reflejando una medida general de la percepción de la innovación educativa.

Como limitación, se debe tomar en cuenta que el tamaño de muestra no permite establecer claramente la diferencia entre el modelo correcto y con problemas de especificación leves, por lo que se deben realizar investigaciones con tamaños de muestras mayores, que permitan detectar problemas de especificación que en estos momentos, no son detectables.

El análisis de confiabilidad indicó que para una medición confiable la escala se debe aplicar al menos a 16 estudiantes, número generalmente inferior a los participantes de las asignaturas en educación superior, ratificando la aplicabilidad de la escala en las aulas universitarias.

Como resultado, se presenta una escala para evaluar innovaciones educativas (EVINED), con evidencia que respalda su validez para evaluar este constructo, de uso amplio para obtener resultados empíricos que permitirá al profesorado, desde la percepción del estudiantado, evaluar la efectividad de nuevas estrategias implementadas en las asignaturas, otorgando información valiosa y confiable para planificar procesos educativos de calidad.

Los resultados de la escala aplicada revelan una visión tradicionalista del proceso enseñanza aprendizaje, a pesar de tratarse de innovaciones pedagógicas, pues se otorga el mayor valor al rol del docente en la implementación, especialmente a su dominio metodológico, además de un valor alto y consistente para todos los ítems relacionados con el aprendizaje disciplinar del estudiantado. A la vez se evidencia el potencial de innovar en el aula, percibiendo la promoción especialmente de la responsabilidad estudiantil por el aprendizaje, probablemente a través de experiencias de aprendizaje activo, que caracterizan a las innovaciones al salir de la tradicional clase conferencia y favorecer la participación estudiantil. Llama la atención que se perciba que la diversidad de diseños promueve en menor medida la comunicación oral, aspecto a cautelar en los diseños de las innovaciones.

A diferencia de otros instrumentos disponibles que se enfocan en el profesorado (Cifuentes y Herrera-Velásquez, 2019; López-Pérez et al., 2023; Santos Rego et al., 2017; Dueñas Zorrilla et al., 2024), EVINED recoge la percepción del estudiantado. Destaca su versatilidad al ser aplicable a cualquier innovación y no a contextos específicos, como la escala de Pozuelos Estrada et al. (2021), incluyendo dimensiones centradas en logros estudiantiles y no principalmente en el rol del docente y del contexto en promover estos logros, como la escala de Traver-Martí y Fernández-Barrueco (2016).

Es esperable que, al contar con instrumentos de medida integrales en sus dimensiones y de amplia aplicación, se promueva la divulgación de experiencias en publicaciones científicas y responder a la brecha reportada de carencia de evaluaciones de procesos

de innovación en el aula (Pedró, 2015). Esta escala responde a la necesidad reportada en instituciones de educación superior de enfatizar en la importancia de proveer oportunidades para que el estudiantado entregue retroalimentación concreta y de calidad (Boden, 2019), permitiendo a los centros de formación del profesorado desarrollar capacidades e instancias de difusión de la cultura de la evidencia en el ejercicio docente (Pedró, 2015). Además, esto puede contribuir al reto de realizar intervenciones innovadoras en red para crear más conexiones entre el profesorado (Stasewitsch et al., 2022). En otras palabras, que se construyan comunidades de aprendizaje, para compartir resultados, con intereses y objetivos comunes, para colaborar en la implementación de innovaciones de calidad. Esto en la medida de que se fomente utilizar instrumentos como este en los procesos de evaluación de las innovaciones educativas. De esta forma, la innovación docente podría ser sostenible en el tiempo a través de la retroalimentación, la interacción, el aprendizaje compartido y la reflexión docente continua (Navareño, 2020).

Una proyección de este trabajo es identificar diseños más efectivos para grupos específicos, al analizar diferencias estadísticamente significativas en los resultados obtenidos en la escala según las variables sociodemográficas e indagar sobre la naturaleza de las innovaciones. Así mismo, se robustecería la comprensión al explorar cualitativamente los ítems con mayor o menor valoración.

## Referencias

- Amante, S., y Fernandes, R. (2023). Aligning HE Pedagogical Innovation with VET, Industry, and Research Partnerships: Insights on the Demola Portugal Initiative. *Education Sciences*, 13(1), art. 93. <https://doi.org/10.3390/educsci13010093>
- Bethhäuser, B. A., Bach-Mortensen, A. M., y Engzell, P. (2023). A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic. *Nature Human Behaviour*, 7(3), 375-385. <https://doi.org/10.1038/s41562-022-01506-4>
- Boden, K. E. (2019). Pedagogical innovation among university faculty. *Creative Education*, 10(5), 848-861. <https://doi.org/10.4236/ce.2019.105063>
- Bruna, C., Espinoza, C., Fernández, C., Sánchez, J., y Díaz, E. (2020). Validez de contenido de instrumento para evaluar percepción de estudiantes sobre intervenciones pedagógicas en aulas universitarias. *X Congreso Internacional sobre Formación en Ciencia, Tecnología y Competencias, CIFCOM2020 «Innovación disruptiva y los nuevos escenarios»* (pp. 208-220). Editorial CIMTED.
- Burganova, R. I., Abdugalina, S. E., y Tuyakova, A. E. (2018). Improving the quality of education through student-centered education. *Series of Social and Human Sciences*, 6(322), 102-104. <https://doi.org/10.32014/2018.2224-5294.40>
- Calma, A., y Dickson-Deane, C. (2020). The student as customer and quality in higher education. *International Journal of Educational Management*, 34(8), 1221-1235. <https://doi.org/10.1108/IJEM-03-2019-0093>
- Chandra, P., Tomitsch, M., y Large, M. (2021). Innovation education programs: a review of definitions, pedagogy, frameworks and evaluation measures. *European Journal of Innovation Management*, 24(4), 1268-1291. <https://doi.org/10.1108/EJIM-02-2020-0043>
- Cifuentes, G. A., y Herrera-Velásquez, D. A. (2019). Scale development and validation to measure institutional conditions to promote educational innovation with ICT. *Education Policy Analysis Archives*, 27(88), 1-19. <https://doi.org/10.14507/epaa.27.3779>
- Cook C., Jones, J., y Al-Twal, A. (2022). Validity and fairness of utilising student evaluation of teaching (SET) as a primary performance measure. *Journal of Further and Higher Education*, 46(2), 172-184. <https://doi.org/10.1080/0309877X.2021.1895093>
- Davis, L. L. (1992). Instrument review: Getting the most from a panel of experts. *Applied Nursing Research*, 5(4), 194-197. [https://doi.org/10.1016/S0897-1897\(05\)80008-4](https://doi.org/10.1016/S0897-1897(05)80008-4)
- den Brok, P. (2018). *Cultivating the growth of life-science graduates: On the role of educational ecosystems. Inaugural lecture*. Wageningen University and Research. <https://doi.org/10.18174/458920>
- Dueñas Zorrilla M., Tejada Fernández, J., y Pozos Pérez K.V. (2024). Diseño y validación de una escala de autoevaluación de la competencia digital docente y actitud hacia la innovación educativa del profesorado en servicio. *Revista Complutense de Educación*, 35(2), 239-252. <https://doi.org/10.5209/rced.85257>
- Eid, M., Geiser, C., Koch, T., y Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods*, 22(3), 541-562. <https://doi.org/10.1037/met0000083>
- Findikoğlu, F., y İlhan, D. (2016). Realization of a Desired Future: Innovation in Education. *Universal Journal of Educational Research*, 4(11), 2574-2580. <https://doi.org/10.13189/ujer.2016.041110>
- Galindo-Domínguez, H., Valero Esteban, J. M., y Verde Trabada, A. (2022). La investigación educativa como elemento clave en el desarrollo de la competencia emprendedora. *Aula Abierta*, 51(1), 17-26. <https://doi.org/10.17811/rife.51.1.2022.17-26>
- Halász, G. (2021). Measuring innovation in education with a special focus on the impact of organisational characteristics. *Hungarian Education Research Journal*, 11(2), 189-209. <https://doi.org/10.1556/063.2021.00032>
- Henríquez Cabezas, N., y Vargas Escobar, D. (2022). Modelos predictivos de rendimiento y deserción académica en estudiantes de primer año de una universidad pública chilena. *Revista de Estudios y Experiencias en Educación*, 21(45), 299-316. <http://dx.doi.org/10.21703/0718-5162.v21.n45.2022.015>
- Hu, L. T., y Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Kauffeld, S., Stasewitsch, E., de Wall, K., y Othmer, J. (2019). Innovationen in der Hochschullehre – das Beispiel Technische Universität Braunschweig. En S. Kauffeld, y J. Othmer (Eds.), *Handbuch Innovative Lehre* (pp. 1–42). Springer. [https://doi.org/10.1007/978-3-658-22797-5\\_1](https://doi.org/10.1007/978-3-658-22797-5_1)
- Kezar, A., y Eckel, P. D. (2002). The effect of institutional culture on change strategies in higher education: Universal principles or culturally responsive concepts? *The Journal of Higher Education*, 73(4), 435-460. <https://doi.org/10.1080/00221546.2002.11777159>

- Krstikj, A., Sosa Godina, J., García Bañuelos, L., González Peña, O. I., Quintero Milián, H. N., Urbina Coronado, P. D., y Vanoye García, A. Y. (2022). Analysis of competency assessment of educational innovation in upper secondary school and higher education: A mapping review. *Sustainability*, 14(13), art. 8089. <https://doi.org/10.3390/su14138089>
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563-575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
- Lin, P. H., Huang, L. R., y Lin, S. H. (2022). Why teaching innovation matters: Evidence from a pre- versus peri-COVID-19 pandemic comparison of student evaluation data. *Frontiers in Psychology*, 13, art. 963953. <https://doi.org/10.3389/fpsyg.2022.963953>
- López, B. (2022). How higher education promotes the integration of sustainable development goals—An experience in the postgraduate curricula. *Sustainability*, 14(4), 2271. <https://doi.org/10.3390/su14042271>
- López-Pérez, G. B., Norzagaray-Benítez, C. C., Sotelo-Castillo, M. A., y Fraijo-Figueroa, J. A. (2023). Adaptación y validación de una escala de uso tecnológico docente universitario. *Sinéctica, Revista Electrónica de Educación*, 61, e1576. [https://doi.org/10.31391/s2007-7033\(2023\)0061-015](https://doi.org/10.31391/s2007-7033(2023)0061-015)
- Lotz-Sisitka, H., Wals, A. E. J., Kronlid, D., y McGarry, D. (2015). Transformative, transgressive social learning: Rethinking higher education pedagogy in times of systemic global dysfunction. *Current Opinion in Environmental Sustainability*, 16, 73-80. <https://doi.org/10.1016/j.cosust.2015.07.018>
- Macanchí Pico, M. L., Orozco Castillo, B. M., y Campoverde Encalada, M. A. (2020). Innovación educativa, pedagógica y didáctica. Concepciones para la práctica en la Educación Superior. *Revista Universidad y Sociedad*, 12(1), 396-403.
- McNeish, D., y Wolf, M. G. (2023). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods*, 28(1), 61-88. <https://doi.org/10.1037/met0000425>
- Marchant, J., González, C., y Fauré, J. (2018). The impact of a university teaching development programme on student approaches to studying and learning experience: Evidence from Chile. *Assessment & Evaluation in Higher Education*, 43(5), 697-709. <https://doi.org/10.1080/02602938.2017.1401041>
- Mazandarani, O., y Troudi, S. (2022). Measures and features of teacher effectiveness evaluation: Perspectives from Iranian EFL lecturers. *Educational Research for Policy and Practice*, 21, 19-42. <https://doi.org/10.1007/s10671-021-09290-0>
- Navareño, P. (2020). Asesoramiento externo a la escuela para la innovación sostenible y la mejora continua, desde las comunidades profesionales de aprendizaje. *Avances en Supervisión Educativa*, 34, art. 7. <https://doi.org/10.23824/ase.v0i34.676>
- Organization for Economic Cooperation and Development [OECD]. (2015). *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities*. OECD. <http://dx.doi.org/10.1787/9789264239012-en>
- Otzen, T., y Manterola, C. (2017). Técnicas de muestreo sobre una población a estudio. *International Journal of Morphology*, 35(1), 227-232. <http://doi.org/10.4067/S0717-95022017000100037>
- Pedro, F. (2015). Las políticas de investigación e innovación en educación: Una perspectiva supranacional. *Bordón. Revista de Pedagogía*, 67(1), 39-56. <http://doi.org/10.13042/Bordon.2015.67103>
- Pozuelos Estrada, F. J., García Prieto, F. J., y Conde Vélez, S. (2021). Evaluar prácticas innovadoras en la enseñanza universitaria. Validación de instrumento. *Educación XX1*, 24(1), 69-91. <http://doi.org/10.5944/educXX1.26300>
- Sánchez-Villena, A. R., de La Fuente-Figueroa, V., y Ventura-León, J. (2021). Modelos factoriales de la Escala de Autoestima de Rosenberg en adolescentes peruanos. *Revista de Psicopatología y Psicología Clínica*, 26(1), 47-55. <https://doi.org/10.5944/rppc.26631>
- Santos Rego, M. A., Jover Olmeda, G., Naval, C., Álvarez Castillo, J. L., Vázquez Verdera, V., y Sotelino Losada, A. (2017). Diseño y validación de un cuestionario sobre práctica docente y actitud del profesorado universitario hacia la innovación (CUPAIN). *Educación XX1*, 20(2), 39-71. <https://doi.org/10.5944/educxx1.19031>
- Sola Fernández, M. (2016). ¿Por qué es necesario innovar? En M. Fernández Navas, y N. Alcaraz Salarirche (Coords.), *Innovación Educativa. Más allá de la ficción* (pp. 41-52). Pirámide.
- Stasewitsch, E., Barthauer, L., y Kauffeld, S. (2022). Knowledge Transfer in a Two-Mode Network Between Higher Education Teachers and Their Innovative Teaching Projects. *Journal of Learning Analytics*, 9(1), 93-110. <https://doi.org/10.18608/jla.2022.7427>
- Traver-Martí, J. A., y Ferrández-Berruero, R. (2016). Construcción y validación de un cuestionario de actitudes hacia la innovación educativa en la universidad. *Perfiles educativos*, 38(151), 86-103.
- Villaruel, V., y Bruna, D. (2014). Reflexiones en torno a las competencias genéricas en educación superior: Un desafío pendiente. *Psicoperspectivas*, 13(1), 22-34. <http://dx.doi.org/10.5027/psicoperspectivas-Vol13-Issue1-fulltext-335>
- Walder, A. M. (2017). Pedagogical Innovation in Canadian higher education: Professors' perspectives on its effects on teaching and learning. *Studies in Educational Evaluation*, 54, 71-82. <https://doi.org/10.1016/j.stueduc.2016.11.001>
- Xia, Y., y Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, 51(1), 409-428. <https://doi.org/10.3758/s13428-018-1055-2>
- Xin, X., Shu-Jiang, Y., Nan, P., ChenXu, D., y Dan, L. (2022). Review on A big data-based innovative knowledge teaching evaluation system in universities. *Journal of Innovation & Knowledge*, 7(3), art. 100197. <https://doi.org/10.1016/j.jik.2022.100197>