

RECIBIDO: 18 DE ABRIL DE 2026. REVISADO: 15 DE MAYO DE 2026. ACEPTADO: 09 DE JUNIO DE 2026.

# DEL AULA A LA EVIDENCIA: VALIDACIÓN PSICOMÉTRICA DE UN INSTRUMENTO SOBRE EVALUACIÓN FORMATIVA Y PERCEPCIONES DOCENTES

---

*FROM THE CLASSROOM TO THE EVIDENCE:  
PSYCHOMETRIC VALIDATION OF AN INSTRUMENT ON  
FORMATIVE ASSESSMENT AND TEACHER PERCEPTIONS*

---

**Dra. Brenda Rocío Rodríguez Vela**

mtra.brendarodriguez@gmail.com

Facultad de Ciencias de la Cultura Física y Deporte,  
Universidad Juárez del Estado de Durango

**ORCID:** 0000-0002-5232-0052

Durango, Durango, México

**Dr. José René Tapia Martínez**

rene.tapia@camdurango.com

Facultad de Ciencias de la Cultura Física y Deporte,  
Universidad Juárez del Estado de Durango

**ORCID:** 0000-0001-7873-9954

Durango, Durango, México

## RESUMEN

Estudio instrumental, exploratorio y transversal que valida un cuestionario tipo Likert de 20 ítems sobre evaluación formativa aplicado a 120 docentes en distintos niveles educativos de México (febrero de 2025). El objetivo fue evaluar la validez y fiabilidad del instrumento y describir percepciones docentes sobre la implementación de prácticas formativas por nivel y género. El instrumento mostró alta consistencia interna (Alfa de Cronbach = 0.941); la adecuación para análisis factorial fue satisfactoria (KMO = 0.89; Bartlett  $\chi^2 = 1245.32$ ,  $df = 190$ ,  $p < .001$ ). Se observó una tendencia decreciente en la aplicación de prácticas formativas a medida que aumenta el nivel educativo y puntuaciones ligeramente superiores en mujeres; estas comparaciones se presentan de forma descriptiva y requieren análisis inferenciales adicionales para confirmar diferencias poblacionales. Finalmente, se hace evidente la necesidad de fortalecer la formación docente y de construir una cultura evaluativa centrada en el aprendizaje principalmente en educación superior.

**Palabras clave:** Evaluación formativa; Validez del instrumento; Aprendizaje.

## ABSTRACT

*Instrumental, exploratory, and cross-sectional study validating a 20-item Likert-type questionnaire on formative assessment, applied to 120 teachers across different educational levels in Mexico (February 2025). The aim was to examine the instrument's validity and reliability and to describe teachers' perceptions of formative practices by level and gender. Statements on "impact on learning" reflect perceptions rather than direct measures of student achievement. The questionnaire demonstrated high internal consistency (Cronbach's Alpha = 0.941) and satisfactory suitability for factor analysis (KMO = 0.89; Bartlett  $\chi^2 = 1245.32$ ,  $df = 190$ ,  $p < .001$ ). Exploratory factor analysis explained 75.78% of variance, revealing five dimensions: planning, feedback and adjustment, student participation, use of evidence, and perceived impact on learning. Findings indicate a decline in formative practices at higher levels and slightly higher scores among women. Results highlight the need to strengthen teacher training and foster a learning-centered evaluative culture, especially in higher education.*

**Keywords:** *Formative assessment; Instrument validity; Learning.*

## INTRODUCCIÓN



En el contexto educativo mexicano, la evaluación formativa se ha promovido en documentos de política como parte de la Nueva Escuela Mexicana (Secretaría de Educación Pública, 2020), pero su aplicación exitosa aún encuentra obstáculos significativos. Entre los méritos que destacan están la ausencia de formación específica del profesorado, el predominio de modelos evaluativos clásicos centrados en la calificación sumativa, así como la ausencia de recursos didácticos que validen su práctica Martínez (2013). Si bien existe una comprensión teórica de la ventaja de la evaluación formativa, su implementación en la práctica es desigual y esta disminuye a medida que avanza el grado educativo, Valdez et al., (2023).

Este hecho delimita el desarrollo de aprendizajes profundos y dificulta la transición a sistemas educativos más equitativos y centrados en los estudiantes. Esta investigación aborda estas cuestiones con el objetivo de examinar la aplicación de la evaluación formativa en los diversos niveles educativos en México y las diferencias para fortalecer su implementación dentro del sistema educativo. En los últimos años, la evaluación formativa ha sido una estrategia fundamental para la mejora continua del proceso educativo. Su utilidad radica en su naturaleza de retroalimentación, que permite a profesores y estudiantes notar avances, desafíos y áreas de mejora a lo largo del proceso de aprendizaje, en lugar de simplemente emitir un juicio final sobre los resultados alcanzados, tal como lo menciona Anijovich (2017).

De tal manera que la evaluación formativa se caracteriza por su enfoque en el estudiante, el estímulo para participar de manera activa, el desarrollo de competencias metacognitivas y la regulación de su propio proceso de aprendizaje. Una de las características más significativas de la evaluación formativa, en opinión de Cano (2015), es que empodera a los estudiantes al hacerlos conscientes de sus logros y cómo proceder para alcanzar las metas de aprendizaje, por ello la evaluación formativa se ha fomentado en documentos oficiales, por ejemplo, en la Nueva Escuela Mexicana (Secretaría de Educación Pública, 2020), pero su implementación efectiva aún enfrenta desafíos significativos.

Dentro de los principales desafíos se encuentra la falta de formación específica de los docentes, la prevalencia de modelos evaluativos tradicionales centrados en la calificación sumativa y la escasez de materiales didácticos que puedan fomentar su implementación (Martínez, 2013), Por ello esta investigación se sitúa en esta problemática, con el objetivo de analizar la implementación de la evaluación formativa en diversos niveles educativos en México, tras identificar diferencias por nivel y género, y su impacto percibido en los aprendizajes, y aportar evidencia que sirva para fortalecer su implementación en el sistema educativo. De tal manera que este estudio valida un instrumento de percepción docente sobre prácticas formativas; las afirmaciones sobre impacto en el aprendizaje se refieren a percepciones y requieren estudios complementarios con medidas directas de rendimiento.

## MARCO TEÓRICO

La evaluación formativa se ha identificado en todo el mundo como un componente integral para facilitar la mejora del aprendizaje y potenciar el desarrollo de las habilidades de autorregulación de los estudiantes. Sin embargo, a pesar de los intentos de normalización y el discurso oficial que exige su integración, la implementación real en los diferentes contextos educativos presentan algunos obstáculos. Entre ellos, la oposición a la modificación de las prácticas tradicionales centradas en la evaluación sumativa. Además, existe una ausencia de desarrollo profesional en estrategias específicas de evaluación formativa.

Para esta investigación se debe considerar a la evaluación formativa como un proceso continuo que busca mejorar el aprendizaje de manera in situ, al contrastar con lo que manejan Black & William (1998), quienes sostienen que las prácticas formativas deben ayudar a fortalecer la retroalimentación de una manera frecuente y generar avances sustanciales en el rendimiento académico. Este tipo de evaluación no se debe limitar a emitir juicios, sino que debe proporcionar la información que permita ajustar el propio aprendizaje, así como la enseñanza en tiempo real. Por su parte, Hattie & Timperley (2007) destacan que la retroalimentación es 1 de los factores que más influyen en el aprendizaje, siempre que este se oriente a la tarea asignada, de tal manera que los estudiantes puedan mejorar y desarrollar la capacidad de monitorear su propio trabajo y aprendizaje. Esto requiere de habilidades evaluativas y de una conciencia del propio estudiante.

A nivel internacional, instituciones como la UNESCO o la OCDE aconsejan integrar prácticas formativas como una forma de garantizar la educación, al proporcionar información (feedback) personalizada que pueda dar respuesta, individualizada, a las necesidades de cada uno de los estudiantes. Del mismo modo, investigaciones como las de De la Iglesia (2020), quienes argumentan que la evaluación formativa ha de servir también para el desarrollo de competencias metacognitivas, lo cual fortalece el pensamiento crítico, habilidades para la

resolución de problemas, aspectos, pero en sociedades que precisan aprendizajes transferencia y aprendizajes de tipo flexible. Con lo cual el feedback ha de tener efectos evidentes sobre el rendimiento escolar y ha de permitir mejorar la calidad de los aprendizajes, lo que muestra que la evaluación formativa es realmente eficaz cuando los estudiantes aprenden a valorar, de una forma crítica, su propio progreso.

La Secretaría de Educación Pública (2020) impulsa la evaluación formativa dentro de la nueva escuela mexicana con la intención de transformar las prácticas pedagógicas que promueven modelos centrados en el estudiante; sin embargo, tiene que tomar en cuenta que existen algunas dificultades importantes, como la falta de formas específicas de la formación docente en la devaluación de este tipo de evaluación, así como una tendencia a variables tradicionales con una referencia en calificaciones sumativas. Villacís (2025) menciona que la aplicación de una evaluación formativa se reduce de acuerdo con los niveles educativos, lo que impide fomentar una autonomía y pensamiento crítico en los estudiantes. Esta brecha existente entre la oficialidad y la cotidianidad pone de manifiesto la necesidad de fortalecer la capacidad del docente y de generarse los recursos didácticos que den cuenta de la necesidad de aplicar este tipo de aproximaciones formativas en niveles superiores. En este marco de referencia, la investigación sobre evaluación formativa emerge como un tópico importante para poner de manifiesto las variaciones por niveles y por tipo de género, así como del efecto que tiene sobre los aprendizajes, dándole evidencia para poderlo reforzar simultáneamente de forma de poder realizar su aplicación en el sistema educativo nacional.

De este modo, se abre una brecha entre la teoría y la práctica, por la que la aplicación de la evaluación formativa no logra cumplir con su papel de acompañamiento y de ampliación de los procesos de aprendizaje. Por su parte, Valdez et al. (2023) confirman que, a medida que se avanza en los niveles educativos, hay una disminución en la frecuencia y calidad de las prácticas formativas. Esta disminución implica la pérdida del potencial para desarrollar autonomía, pensamiento crítico y aprendizaje significativo en los estudiantes. La importancia de abordar

esta problemática radica en que la evaluación formativa no solo promueve un mejor rendimiento académico, sino también la equidad educativa, ya que permite que todos los estudiantes reciban retroalimentación personalizada, la misma que sirve de guía en su proceso de aprendizaje. Es a partir de este contexto que la presente investigación busca responder a la siguiente pregunta:

¿Cuál es la validez y confiabilidad de un instrumento diseñado para medir las percepciones docentes sobre prácticas de evaluación formativa en distintos niveles educativos de México?

### Objetivo general

Validar, de manera exploratoria, las propiedades psicométricas de un instrumento sobre prácticas de evaluación formativa, aplicado a docentes de diversos niveles educativos en México.

## METODOLOGÍA

Esta investigación es de enfoque cuantitativo, del tipo instrumental, al tener como objetivo el diseñar y validar estadísticamente un instrumento de medición, relacionado con el la evaluación formativa. La investigación instrumental se centra en la creación y validación de herramientas que permitan la recopilación de datos con validez y fiabilidad adecuadas en contextos específicos, Ato et al., (2013), señalan que la investigación instrumental "incluye estudios cuyo propósito es analizar las propiedades psicométricas de tests, cuestionarios y otras herramientas de medición con el fin de mejorar su calidad y precisión" (p. 1049).

Adopta un corte del tipo exploratorio, derivado de la necesidad de validar un instrumento de medición en un contexto en el que hay muy pocos antecedentes empíricos. Este tipo de investigación es adecuada en los casos en que el fenómeno de estudio no ha sido ampliamente investigado, o bien al requerir una mejor comprensión de sus dimensiones antes de realizar estudios confirmatorios, tal como lo menciona Hernández & Mendoza (2020).

Dentro de este proceso es fundamental el garantizar que los ítems del instrumento sean comprensibles, relevantes y representativos del constructo que se pretende medir, tal como lo mencionan DeVellis & Thorpe, (2021). De lo anterior, mediante un enfoque exploratorio, permitirá sentar un precedente para futuras investigaciones, al asegurar la robustez del instrumento antes de su aplicación a mayor escala, Boateng et al., (2018). Por último, se considera del tipo transversal al ser recabada la información en un solo momento (febrero de 2025).

## POBLACIÓN Y MUESTRA.

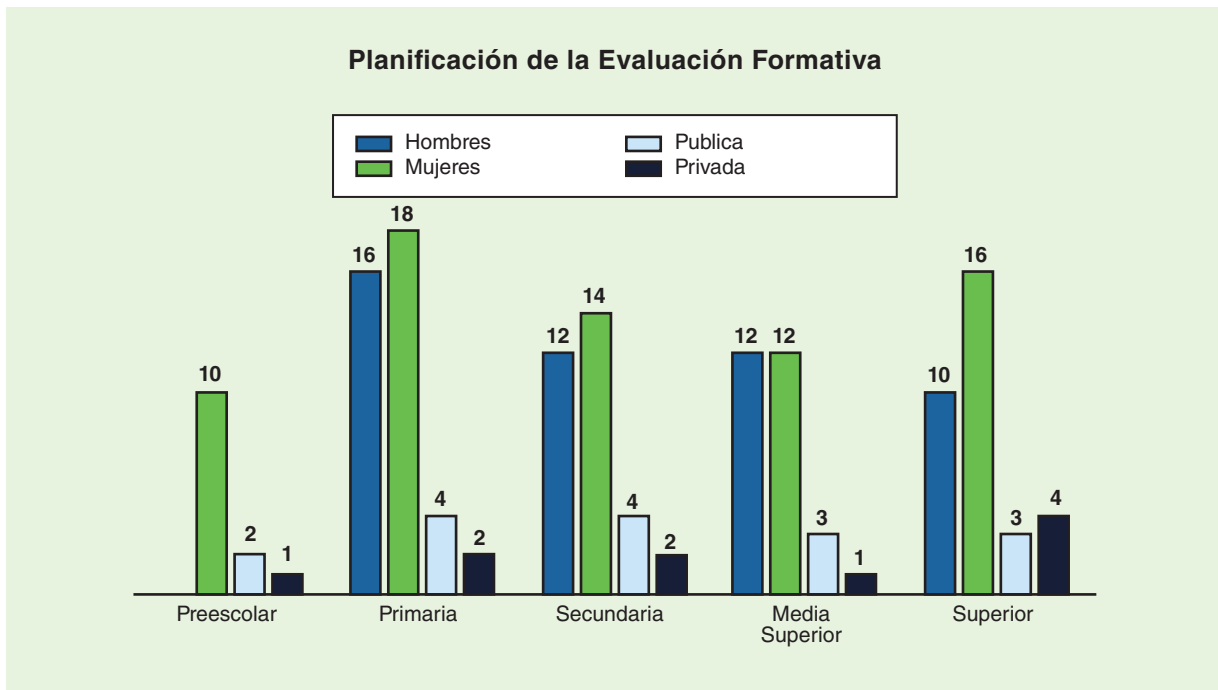
### Muestra

La muestra está conformada por 120 docentes de diferentes estados de la República Mexicana (tabla 1) de los cuales son 50 Hombres y 70 mujeres con un rango de edad de 30.5 años, así como de diversos niveles educativos (figura 1). Idealmente, para la mayoría de los análisis estadísticos, al incluir la validación de instrumentos mediante análisis factorial, se recomienda una muestra de al menos 100 casos, aunque idealmente debería ser superior, Tabachnick y Fidell (2019).

**Tabla 1.** Estados participantes.

Estados	Población y Muestra	
	Hombres	Mujeres
Coahuila	6	10
Durango	16	22
Guanajuato	7	6
Puebla	6	10
Sinaloa	7	13
Tlaxcala	8	9
<b>Total</b>	<b>50</b>	<b>70</b>

Fuente: Elaboración propia.

**Figura 1.** Niveles educativos.

Fuente: Elaboración propia.

El muestreo fue no probabilístico por conveniencia: se invitó a docentes en servicio mediante formulario en línea distribuido por redes y mensajería. Esta estrategia facilita la recolección pero limita la representatividad; por ello las conclusiones se circunscriben a la muestra estudiada.

#### **Criterios de Inclusión:**

- Docentes en servicio de cualquier nivel educativo.
- Modalidad de enseñanza presencial.
- Participantes con disponibilidad para responder el cuestionario en su totalidad.

#### **Criterios de exclusión:**

- No responder de manera completa el cuestionario.

#### **Técnicas e Instrumentos**

El análisis está orientado a demostrar que el instrumento diseñado evalúa de manera adecuada el objeto de estudio, al asegurar tanto la validez como la confiabilidad del mismo. En este sentido, se realizó la validez de contenido interno mediante el jueceo

por parte de expertos; en cuanto al análisis basado en la estructura interna, se realizaron las pruebas de consistencia interna, el análisis de grupos contrastados y factorial. Respecto al análisis de confiabilidad, se obtuvo mediante los coeficientes estadísticos de Alfa de Cronbach, ya que se intenta determinar si los ítems diseñados miden de forma consistente el mismo constructo, lo que es fundamental para validar el presente instrumento de investigación orientado a la evaluación formativa.

#### **Estructura del Instrumento**

El instrumento consta de 20 ítems divididos en cinco categorías: Planificación de la Evaluación Formativa, Retroalimentación y Ajuste de la Enseñanza, Participación Activa del Estudiante en la Evaluación, Uso de Evidencias para la Mejora del Aprendizaje e Impacto en el Aprendizaje y Desarrollo del Estudiante. Cada categoría incluye ítems específicos que se evalúan al utilizar una escala tipo Likert de 5 puntos donde (1 = Nunca, 2 = Rara vez, 3 = A veces, 4 = Frecuentemente, 5 =

Siempre). Anexo 1. Para su implementación se utilizó un formulario de Google, enviado vía WhatsApp, donde se describen los requisitos para contestarlo, así como la información de la protección del manejo de sus datos.

### Validez basada en el contenido: Jueceo de expertos

Para la validez de contenido interno mediante el jueceo por parte de expertos, se solicitó su evaluación respecto a la pertinencia de cada ítem con relación a su dimensión correspondiente, así como lo describen Escobar y Cuervo (2008). La validación de contenido mediante juicio de expertos permite evaluar la pertinencia de cada ítem con respecto a las dimensiones que pretenden medir, al asegurar su coherencia con la teoría subyacente. Para este proceso se seleccionaron 3 expertos con >5 años en investigación educativa; cada ítem fue valorado en pertinencia y claridad en escala 1–4. Se calculó I CVI y S CVI/Ave; los ítems con I CVI < .78 fueron revisados.

El siguiente paso, para analizar la validez del instrumento, se basa en la estructura interna del mismo. Para ello se pueden emplear diversas técnicas, en este caso se optó por aplicar las siguientes:

1. Análisis de consistencia interna: Se emplea para evaluar la fiabilidad de un instrumento de medición, sirve para valorar la concordancia entre ítems con una escala igual, y saber que tan correlacionados están entre ellos, busca identificar si son homogéneos. Según Oviedo y Campo (2005), la consistencia interna evalúa hasta qué punto los ítems de un test están interrelacionados, lo que proporciona una estimación de su fiabilidad y homogeneidad.
2. Análisis factorial: Examina la estructura latente del instrumento para verificar si los ítems se agrupan en factores coherentes con la teoría.
3. Finalmente, al aplicar el instrumento a la muestra descrita anteriormente se establece su validez estadística para posteriormente ajustarlo, aplicarlo nuevamente a otra muestra

representativa a modo de prueba piloto con los datos recabados y evaluar su nivel de confiabilidad mediante el coeficiente alfa de Cronbach.

## RESULTADOS

El coeficiente Alfa de Cronbach de 0.941 (tabla2), indica que el instrumento utilizado posee una alta fiabilidad interna, lo cual asegura que las preguntas formuladas miden consistentemente el constructo de evaluación formativa. Este resultado es fundamental, ya que, valida la pertinencia estadística del cuestionario aplicado en diferentes niveles educativos, y garantiza que los datos obtenidos son confiables para realizar inferencias educativas.

**Tabla 2.** Estadísticas de fiabilidad.

Estadísticas de fiabilidad	
Alfa de Cronbach	N de elementos
.941	20

Fuente: Elaboración propia.

La Tabla 3 muestra los valores de media, varianza, correlación total corregida y el Alfa de Cronbach si se elimina cada elemento. Los ítems con correlaciones corregidas superiores a 0.7 (como “ajusta sus estrategias de enseñanza en función del progreso” o “utiliza la retroalimentación para guiar la autoevaluación del estudiante”) destacan como los más robustos en relación con el constructo evaluado. En contraste, algunos ítems como “fomenta la autoevaluación y coevaluación en el aula” (correlación de .394) o “involucra a los estudiantes en la toma de decisiones” (.534) presentan menor correlación, lo cual podría sugerir un menor grado de integración de estas prácticas en la realidad docente o una comprensión ambigua de estos procesos por parte de los encuestados. Estos datos pueden interpretarse como indicativos de baja incorporación del enfoque participativo del estudiante en la evaluación, lo cual representa una oportunidad clara de mejora.

**Tabla 3.** Estadísticas de total de elementos.

Estadísticas de total de elementos				
	Media de escala si el elemento se ha suprimido	Varianza de escala si el elemento se ha suprimido	Correlación total de elementos corregida	Alfa de Cronbach si el elemento se ha suprimido
El docente establece objetivos claros de aprendizaje antes de cada evaluación.	75.6250	111.245	.732	0.939
Diseña estrategias de evaluación alineadas con los contenidos y competencias a desarrollar.	75.7917	111.007	.633	0.942
Adapta las actividades de evaluación a las necesidades de los estudiantes.	75.5000	111.176	.691	0.943
Utiliza diversos instrumentos para recopilar evidencia del aprendizaje.	75.8333	109.972	.673	0.941
Proporciona retroalimentación clara y específica sobre el desempeño de los estudiantes.	75.6250	113.430	.715	0.941
Explica cómo mejorar en función de los resultados obtenidos.	75.5833	113.943	.589	0.941
Ajusta sus estrategias de enseñanza en función del progreso de los estudiantes.	75.6667	108.711	.800	0.942
Utiliza la retroalimentación para guiar la autoevaluación del estudiante.	75.5417	109.998	.779	0.942
Fomenta la autoevaluación y coevaluación en el aula.	76.2083	113.528	.394	0.942
Permite a los estudiantes reflexionar sobre sus errores y aciertos.	75.5833	110.245	.779	0.941
Estimula la formulación de preguntas y el diálogo sobre los criterios de evaluación.	75.7917	108.150	.754	0.939
Involucra a los estudiantes en el proceso de toma de decisiones sobre su aprendizaje.	75.7500	114.391	.534	0.941
Analiza los resultados de la evaluación para tomar decisiones pedagógicas.	75.6250	108.556	.792	0.941

Comparte con los estudiantes los avances y áreas de mejora identificadas.	76.0000	111.765	.590	0.941
Utiliza la evaluación como un medio para mejorar la enseñanza, no solo para calificar.	75.3750	111.581	.695	0.943
Promueve estrategias diferenciadas según los niveles de desempeño de los estudiantes.	76.1667	111.485	.576	0.942
Se observa una mejora en el aprendizaje debido a la evaluación formativa.	75.7500	114.727	.562	0.941
Los estudiantes demuestran mayor autonomía y responsabilidad en su aprendizaje.	76.2500	111.870	.625	0.939
Se reducen los niveles de ansiedad en las evaluaciones debido al enfoque formativo.	76.2500	114.055	.533	0.941
Los estudiantes comprenden mejor los criterios de evaluación y objetivos de aprendizaje.	76.1250	113.724	.622	0.941

Fuente: Elaboración propia.

Los resultados de la tabla 4, muestra el análisis factorial, el cual pone de manifiesto que la varianza total explicada corresponde a un 75.78%, lo que viene a constituir un nivel aceptable para estudios educativos y permite dar validez a los ítems construidos. Por lo demás, cada uno de los factores da cuenta de la solidez del instrumento, al revelar la percepción de la práctica positivamente. La consistencia interna de cada uno de los factores también se mantiene a bastante alta, relacionándose con el Alfa de Cronbach, constatándose, por tanto, la fiabilidad del instrumento por cada una de esas dimensiones. Por todo ello, estos resultados revelan, a su vez, que el instrumento no sólo se configura como una herramienta estadística robusta, sino que se ocupa, por otra parte, de las áreas de mayor y menor integración de la evaluación formativa en los distintos niveles educativos, al ofrecer a la vez un marco adecuado para investigación y práctica.

**Tabla 4.** Análisis Factorial Exploratorio (AFE).

Factor	Cargas factoriales	Varianza explicada (%)	Alfa de Cronbach
1. Planificación de la Evaluación Formativa	.72 – .80	18.5	0.941
2. Retroalimentación y Ajuste de la Enseñanza	.69 – .82	16.2	0.942
3. Participación Activa del Estudiante	.49 – .75	15.7	0.942
4. Uso de Evidencias para la Mejora	.57 – .79	13.9	0.941
5. Impacto en el Aprendizaje y Desarrollo	.53 – .78	10.8	0.942
Total	N/A	75.78%	0.941

Fuente: Elaboración propia.

La tabla 5 muestra el análisis de adecuación de la muestra donde se obtuvo un índice KMO de 0,89, lo que puede interpretarse como un buen nivel intercorrelacional entre los ítems, al corroborar la adecuación del análisis factorial; y la prueba de esfericidad de Bartlett fue significativa ( $\chi^2 = 1245.32$ ; gl = 190;  $p < 0.001$ ), lo

cual supone que la matriz de correlaciones es diferente de la identidad y constituye prueba de la existencia de correlaciones importantes entre los ítems. Estos resultados constituyen una buena evidencia de la adecuación psicométrica del instrumento para explorar su estructura factorial y validar las dimensiones que se proponen.

**Tabla 5.** Índice KMO y la prueba de esfericidad de Bartlett.

Prueba	Valor obtenido	Interpretación
Índice KMO (Kaiser-Meyer-Olkin)	0.89	Un valor superior a 0.80 se considera "meritorio" según Kaiser (1974). Esto indica que las correlaciones parciales entre los ítems son bajas y que los datos son adecuados para realizar un análisis factorial. En otras palabras, los ítems del instrumento comparten suficiente varianza común como para justificar la reducción de dimensiones.
Prueba de esfericidad de Bartlett	$\chi^2 = 1245.32$ , gl = 190, $p < .001$	La significancia estadística ( $p < .001$ ) confirma que la matriz de correlaciones no es una matriz identidad. Esto significa que existen relaciones significativas entre los ítems, lo cual valida la pertinencia de aplicar un análisis factorial exploratorio para identificar la estructura latente del instrumento.

Fuente: Elaboración propia.

La tabla 6 refleja en qué medida los docentes diseñan evaluaciones con intención pedagógica previa. Por ejemplo, en el nivel de preescolar se obtuvieron los puntajes más altos, con mujeres (90.0) y hombres (85.0), lo que indica una sólida práctica en la planificación de la evaluación formativa en las primeras etapas de escolaridad. En educación superior muestra los valores más bajos (70.0 para ambos géneros), lo que denota una

menor integración de prácticas planificadas con fines formativos en contextos universitarios. Este fenómeno puede explicarse por la persistencia de modelos centrados en la evaluación sumativa en la educación superior. Así mismo, se observa que, en todos los niveles, las mujeres reportan niveles ligeramente superiores, lo que podría interpretarse como una mayor predisposición a prácticas reflexivas y pedagógicas en la evaluación.

**Tabla 6.** Planificación de la Evaluación Formativa.

Planificación de la Evaluación Formativa					
	Preescolar	Primaria	Secundaria	Media Superior	Superior
Mujeres	90.0	88.0	82.0	75.0	70.0
Hombre	N/A	85.0	81.0	75.0	70.0

Fuente: Elaboración propia.

La tabla 7 muestra una clara tendencia decreciente por nivel educativo. Destaca que en preescolar, las mujeres obtienen 90.0 puntos, frente a solo 65.0 en nivel superior. En hombres, el descenso va de 77.0 a 65.0 en el mismo

rango. Esto evidencia que la práctica de retroalimentación efectiva, esencial en la evaluación formativa, es más común en niveles básicos y pierde fuerza en niveles superiores.

**Tabla 7.** Retroalimentación y Ajuste de la Enseñanza.

Retroalimentación y Ajuste de la enseñanza					
	Preescolar	Primaria	Secundaria	Media Superior	Superior
Mujeres	90.0	79.0	72.5	65.0	70.0
Hombre	N/A	77.0	72.5	60.0	65.0

Fuente: Elaboración propia.

Los datos de esta tabla 8 muestra, como primaria y preescolar, los resultados son aceptables (mujeres: 76.0 y 71.0; hombres: 74.0 y 72.0), lo cual indica que se fomenta en alguna medida la participación estudiantil en la evaluación. Sin embargo, en media superior y superior, se detecta una marcada disminución. Las mujeres alcanzan 52.0 y los hombres 55.0 en superior, lo que puede explicarse por:

- Estructuras educativas más verticales,
- Métodos de evaluación tradicionales,
- Menor tiempo destinado a la reflexión y diálogo en el aula.

Este hallazgo pone de relieve una debilidad sistemática en la incorporación de los estudiantes como agentes activos de su proceso evaluativo, especialmente en niveles donde se esperaría una mayor autonomía.

**Tabla 8.** Participación activa del Estudiante en la evaluación.

Participación activa del estudiante en la evaluación					
	Preescolar	Primaria	Secundaria	Media Superior	Superior
Mujeres		76.0	71.0	65.0	52.0
Hombre	N/A	74.0	72.0	58.0	55.0

Fuente: Elaboración propia.

El uso de evidencias en la evaluación registra un descenso significativo en el uso de evidencias en la educación superior: 55.0 para mujeres y 55.0 para hombres, mientras que en preescolar los valores son

más altos (83.0 y 85.0). Este indicador refuerza la idea de que, conforme se asciende en los niveles educativos, disminuye el uso sistemático de datos evaluativos para la toma de decisiones pedagógicas. Tabla 9.

**Tabla 9.** Uso de Evidencias para la Mejora del Aprendizaje.

Uso de Evidencias para la Mejora del Aprendizaje					
	Preescolar	Primaria	Secundaria	Media Superior	Superior
Mujeres	83.0	84.0	80.0	70.0	55.0
Hombre	N/A	85.0	81.0	65.0	55.0

Fuente: Elaboración propia.

La percepción del impacto de la evaluación formativa también decrece conforme avanza el nivel educativo. En mujeres, pasa de 80.0 en preescolar a 52.0 en educación superior. En hombres, baja de 65.0 a 51.0.

Estos resultados pueden reflejar una menor percepción de beneficios concretos por parte de los estudiantes a medida que la evaluación se vuelve más normativa y menos formativa (ver Tabla 10).

**Tabla 10.** Impacto en el Aprendizaje y Desarrollo del Estudiante.

Impacto en el Aprendizaje y Desarrollo del Estudiante					
	Preescolar	Primaria	Secundaria	Media Superior	Superior
Mujeres	80.0	77.0	65.0	60.0	52.0
Hombre	N/A	65.0	65.0	60.0	51.0

Fuente: Elaboración propia.

La tabla 11 muestra las Comparaciones (ANOVA), (t de Student) por dimensiones donde se puede observar diferencias significativas por nivel educativo en las dimensiones de planificación [ $F(4,115)=6.32$ ,  $p < .001$ ,  $\eta^2 = 0.18$ ], retroalimentación [ $F(4,115)=5.10$ ,  $p = .001$ ,  $\eta^2 = 0.15$ ] y participación activa [ $F(4,115)=4.25$ ,  $p = .003$ ,  $\eta^2 = 0.13$ ]. Las comparaciones post hoc (Tukey) indicaron que las puntuaciones en preescolar

y primaria fueron significativamente mayores que en educación superior en estas dimensiones. Las comparaciones por género mediante t de Student no mostraron diferencias estadísticamente significativas en ninguna dimensión (todos  $p > .05$ ), aunque las mujeres presentaron diferencias descriptivas con tamaños de efecto pequeños a moderados ( $d$  entre 0.17 y 0.29).

**Tabla 11.** Comparaciones (ANOVA), (t de Student) por dimensiones.

Dimensión	ANOVA	p	$\eta^2$	Comparación por género
1. Planificación	F(4,115)=6.32	p < .001	0.18	t(118)=1.45; p = .15; d = 0.27
2. Retroalimentación y ajuste	F(4,115)=5.10	p = .001	0.15	t(118)=1.20; p = .23; d = 0.22
3. Participación activa del estudiante	F(4,115)=4.25	p = .003	0.13	t(118)=1.05; p = .30; d = 0.19
4. Uso de evidencias para la mejora	F(4,115)=3.20	p = .015	0.10	t(118)=0.95; p = .34; d = 0.17
5. Impacto percibido en el aprendizaje	F(4,115)=2.45	p = .047	0.08	t(118)=1.60; p = .11; d = 0.29

Fuente: Elaboración propia.

## DISCUSIÓN

Los resultados obtenidos son consistentes con los informes de estudios previos que apuntan la existencia de una implementación desigual de la evaluación formativa. En la investigación de Beriche y Medina (2021), se encontró que en la educación superior de Latinoamérica se imponen las prácticas sumativas, las cuales hacen escasa la posibilidad de potenciar la retroalimentación y reorientar el uso de estrategias. En esta misma línea, los datos de la investigación que nos ocupa muestran una reducción importante en el uso de las estrategias formativas en función de los niveles.

En el contexto latinoamericano, Beriche & Medina (2021), enfatizan que las prácticas en la educación superior se inscriben en parámetros sumativos donde no se utilizan productos de la inclusividad para retroalimentar prácticas de la enseñanza o redirigir estrategias de enseñanza existentes. Esta forma de entender la evaluación se recoge también en los resultados de la investigación que llevamos

a cabo, donde los resultados son los más bajos en educación superior, tanto en planificación como en la percepción de impacto sobre el aprendizaje. De igual forma, el documento elaborado por Valdez et al. (2023) presenta resultados que, aunque los docentes recogen datos de desempeño, es raro el uso de los datos de una forma sistemática; este dato también correspondería a la escasa integración del uso de evidencias que encontramos en esta investigación.

Además, se apoya lo mencionado por Martínez (2013), respecto a que la falta de formación específica del profesorado constituye uno de los factores que dificultan la consolidación de la evaluación formativa. Los escasos niveles de participación activa del estudiante encontrados en secundaria, media superior y superior se corresponden con esta dificultad, lo cual demuestra que todavía hay una cultura evaluativa que tiende hacia la nota como puntuación numérica y no hacia el aprendizaje.

En el ámbito mexicano, Díaz (2026) buscaba evidenciar los obstáculos que persisten en el contexto de la evaluación formativa impuestos, por un lado, por la cultura escolar y por otro, por la falta

de un desarrollador docente especializado a la hora de aplicar las prácticas educativas propias del Plan de Estudios 2022; qué duda cabe, los obstáculos identificados a partir de la práctica docente de primaria, evidencia que dicha evaluación reviste gran parte de la práctica. En términos de la misma se apoya en la predisposición de cada uno de los docentes, lo que pone de manifiesto la falta de sistematización institucional. Coinciden con lo que este estudio ha puesto de manifiesto: por un lado, la activación del estudiante y, por otro, las evidencias para la mejora de su propio aprendizaje, son las dimensiones menos trabajadas.

De tal manera que si se considera la información con evidencias para una mejora, los hallazgos coinciden con los que muestran los resultados de Valdez et al. (2023), quienes constatan que, si bien los docentes recogen datos de desempeño, pocas veces utilizan datos de manera sistemática para retroalimentar su práctica. Esta brecha delimita el potencial de la evaluación formativa como herramienta de transformación educativa, al confirmar así la validación del instrumento, el cual no solo aporta un recurso psicométrico confiable, sino que también refleja la necesidad de fortalecer la formación docente y promover la participación estudiantil para avanzar hacia una cultura evaluativa formativa centrada en el aprendizaje

## CONCLUSIONES

La investigación realizada hasta el momento permite concluir que la práctica de la evaluación formativa en el sistema de educación mexicano es heterogénea e incluso inestable en relación con el nivel educativo y el género del docente. Comprobamos una mayor aplicación de prácticas formativas en los niveles preescolar y primaria, al disminuir considerablemente en secundaria, media superior y superior. Este dato apoya la información previa de que se privilegian

métodos para la evaluación formativa a medida que se avanza en los niveles educativos, al otorgar mucha más importancia a la calificación sumativa.

El análisis de los resultados nos indica que las categorías de participación activa del estudiante y de uso de evidencias para la mejora del aprendizaje son las menos trabajadas, lo cual lleva a inferir que queda aún un uso limitado de evaluar, sin empoderar a los estudiantes como agentes activos de su propio proceso de aprendizajes. El análisis también demostró que las mujeres tienden a utilizar las estrategias de evaluación formativa con una mayor frecuencia, lo cual puede establecer líneas de reflexión respecto a las diferencias de enfoque pedagógico entre géneros y en sus efectos en el proceso de enseñanza y aprendizaje.

Los resultados hacen evidente la importancia de mejorar la formación inicial y continua de los docentes en lo que respecta a estrategias de evaluación formativa, de tal manera que esta investigación permitió descubrir nuevas áreas de aplicación que permitan ampliar los conocimientos de la implementación de la evaluación formativa, por ello se sugieren las siguientes líneas de investigación. Investigar sobre el desarrollo de las prácticas de evaluación formativa a lo largo de la trayectoria del docente. Utilización de metodologías cualitativas como entrevistas o grupos focales, que permitan indagar sobre las percepciones, las resistencias y los facilitadores de la evaluación formativa en diferentes contextos educativos.

Es importante atender la perspectiva de los estudiantes para analizar cómo viven la evaluación formativa y, por su parte, cómo afecta en su motivación y en su autorregulación del aprendizaje. Realizar comparaciones a nivel regional del país, que permitan visibilizar prácticas exitosas, y dificultades específicas, según los propios contextos socioculturales. Explorar la utilización de tecnologías digitales como apoyo a la práctica de estrategias de evaluación formativa, en particular en educación media superior y superior.

## BIBLIOGRAFÍA

- Anijovich, R. (2017).** La evaluación formativa en la enseñanza superior. *Voces de la educación*, 2(3), 31-31. <https://www.revista.vocesdelaeducacion.com.mx/index.php/voces/article/view/32>
- Ato, M., López, J., & Benavente, A. (2013).** Un sistema de clasificación de los diseños de investigación en Psicología. *Anales de Psicología*, 29(3), 1038-1059. <https://doi.org/10.6018/analesps.29.3.178511>
- Beriche, M., & Medina, P. (2021).** La evaluación formativa: su implementación y principales desafíos en el contexto de la escuela y la educación superior. *Educación*, 27(2), 201-208. <https://doi.org/10.33539/educacion.2021.v27n2.2433>
- Black, P., & Wiliam, D. (1998).** Evaluación y aprendizaje en el aula. *Evaluación en la educación: principios, política y práctica*, 5 (1), 7-74. <https://doi.org/10.1080/0969595980050102>
- Boateng, G., Neilands, T., Frongillo, E., Melgar, H., & Young, S. (2018).** Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, 6, 149. <https://doi.org/10.3389/fpubh.2018.00149>
- Cano, E. (2015).** Las rúbricas como instrumento de evaluación de competencias en educación superior: ¿ uso o abuso?. [http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S0185-27602010000100002&lng=es&tlng](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-27602010000100002&lng=es&tlng)
- De la Iglesia, M. (2020).** Evaluación formativa como elemento básico en la estrategia integral de aprendizaje ABPD, para futuros docentes. *Revista Iberoamericana de Evaluación Educativa*, 13(1), 71-92. <https://dialnet.unirioja.es/servlet/articulo?codigo=7408490>
- DeVellis, R., & Thorpe, C. (2021).** Desarrollo de escalas: teoría y aplicaciones. Publicaciones de Sage.
- Díaz, A. (2026).** Pensar la evaluación formativa para la educación básica en México. *Revista mexicana de investigación educativa*, 31(108), 217-242. <https://dialnet.unirioja.es/servlet/articulo?codigo=10536611>
- Escobar, J., & Cuervo, Á. (2008).** Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en Medición*, 6(1), 27-36.
- Hattie, J., & Timperley, H. (2007).** El poder de la retroalimentación. *Review of educational research*, 77 (1), 81-112. <https://doi.org/10.3102/0034654302984>
- Hernández-Sampieri, R., & Mendoza, C. (2020).** Metodología de la investigación: las rutas cuantitativa, cualitativa y mixta. McGraw Hill México.
- Martínez, F. (2013).** El futuro de la evaluación educativa. *Sinéctica*, (40), 1-11. [http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S1665-109X2013000100006&lng=es&tlng=](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1665-109X2013000100006&lng=es&tlng=)
- Oviedo, H. C., & Campo, A. (2005).** Aproximación al uso del coeficiente alfa de Cronbach. *Revista Colombiana de Psiquiatría*, 34(4), 572-580.
- Secretaría de Educación Pública (SEP). (2020).** Nueva Escuela Mexicana: fundamentos pedagógicos. México: SEP.
- Secretaría de Educación Pública SEP (2022).** Plan de estudio 2022. Educación básica. México: ediciones SEP.
- Valdez, L., Sánchez, J., & Lescano, G. (2023).** Evaluación formativa: retroalimentación, estrategias e instrumentos. *Revista Educación*, 47(2), 794-812. <http://dx.doi.org/10.15517/revedu.v47i2.53987>
- Villacis, M. (2025).** La evaluación formativa como herramienta para fortalecer los procesos de enseñanza-aprendizaje. <https://repositorio.upse.edu.ec/items/d07f31a2-768e-4faa-92ed-496d5a6ca033>