

# Evaluación formativa y calidad de la retroalimentación: diseño y validación de escalas para docentes escolares

## *Formative evaluation and quality of feedback: design and validation of scales for school teachers*

Juan Romeo Dávila Ramírez <sup>1\*</sup> 

Juan Antonio Huertas Martínez <sup>2</sup> 

<sup>1</sup> Universidad de Tarapacá, Chile

<sup>2</sup> Universidad Autónoma de Madrid, España

\* Autor de correspondencia. E-mail: [jdavilar@academicos.uta.cl](mailto:jdavilar@academicos.uta.cl)

### Cómo referenciar este artículo/ How to reference this article:

Dávila Ramírez, J. R., & Huertas Martínez, J. A. (2024). Evaluación formativa y calidad de la retroalimentación: diseño y validación de escalas para docentes escolares [Formative evaluation and quality of feedback: design and validation of scales for school teachers]. *Educación XX1*, 27(2), 167-194. <https://doi.org/10.5944/educxx1.38283>

**Fecha de recepción:** 08/09/2023

**Fecha de aceptación:** 23/01/2024

**Publicado online:** 28/06/2024

### RESUMEN

Se ha descrito a la evaluación formativa como una de las principales prácticas pedagógicas para promover el desarrollo de los aprendizajes, ya que permite que docentes y estudiantes visualicen las brechas entre el nivel de dominio alcanzado y el nivel de dominio esperado, y direccionen sus acciones hacia la mejora. Pese a ello, la operativización de las estrategias que conforman dicha metodología aún es confusa y no se ha definido con precisión qué implica un uso óptimo en cuanto a su calidad. El presente estudio tiene como objetivo diseñar y validar una Escala de Estrategias para la Evaluación Formativa y una Escala de Calidad de la Retroalimentación para el Aprendizaje. Los datos de 364 docentes de primaria y secundaria han sido analizados a través de análisis factoriales exploratorios y análisis

factoriales confirmatorios con modelos de ecuaciones estructurales. Luego se ha realizado un análisis de validación cruzada para cada escala entre dos submuestras aleatorias. También se han explorado las diferencias según el nivel educativo donde se desempeñaba el profesorado participante, primaria o secundaria, y las diferencias según su género. Los resultados han señalado una adecuada bondad de ajuste para la Escala de Estrategias de Evaluación Formativa:  $\chi^2 / gl = 3.2$ , CFI = .91 y RMSEA = .07 y para la Escala de Calidad de la Retroalimentación para el Aprendizaje  $\chi^2 / gl = 1.8$ , CFI = .94 y RMSEA = .05. No se han encontrado diferencias significativas según nivel educativo ni según género. La discusión presenta un modelo heurístico que ilustra las relaciones entre como se ha definido evaluación formativa y calidad de la retroalimentación con procesos del aprendizaje y de la enseñanza, considerando las influencias que ejercen los climas de aula, los climas escolares y los sistemas educativos nacionales.

**Palabras clave:** evaluación formativa, retroalimentación, coevaluación, autoevaluación

## ABSTRACT

Formative evaluation has been described as one of the main pedagogical practices to promote the development of learning since it allows teachers and students to visualize the gaps between the level of mastery achieved and the level of mastery expected and direct their actions towards improvement. Despite this, the operationalization of the strategies that make up said methodology is still confusing and it has not been precisely defined what optimal use implies in terms of quality. The objective of this study is to design and validate a Strategies Scale for Formative Evaluation and a Feedback Quality Scale for Learning. The data of 364 primary and secondary teachers have been analyzed through exploratory factor analysis and confirmatory factor analysis with structural equation models. Then, a cross-validation analysis has been carried out for each scale between two random subsamples. Also, differences have been explored according to the educational level where the participating teachers worked, primary or secondary, and differences according to their gender. The results have indicated an adequate goodness of fit for the Formative Evaluation Strategies Scale:  $\chi^2 / df = 3.2$ , CFI = .91 and RMSEA = .07 and for the Quality of Feedback for Learning Scale  $\chi^2 / df = 1.8$ , CFI = .94 and RMSEA = .05. No significant differences were found according to educational level or gender. The discussion presents a heuristic model that illustrates the relationships between how formative evaluation and quality of feedback have been defined with learning and teaching processes, considering the influences exerted by classroom climates, school climates and national educational systems.

**Keywords:** formative evaluation, feedback, co-evaluation, self-evaluation

## INTRODUCCIÓN

### Aprendizaje y evaluación formativa

Se ha definido al aprendizaje humano como un metaproceto de transformación psicológica que emerge de la interacción entre una persona y un entorno social enriquecedor, como resultado de procesos internos de explicitación e implicitación, que incluyen la integración de contenidos conceptuales y motivacionales (Illeris, 2014; Pozo, 2017). Esto es coherente con la noción de desarrollo próximo que enfatiza el papel de un apoyo experto en el aprendizaje (Vygotsky, 1979). Ambos argumentos relevan el rol de los procesos intersubjetivos en el aprendizaje, que en el caso de la educación formal ocurren entre el estudiantado y el profesorado (Greene et al., 2004). En las escuelas, una de las consecuencias de estas interacciones son las informaciones que surgen sobre la calidad de los aprendizajes, o sea, la retroalimentación, un proceso constitutivo de la evaluación formativa.

En general, se ha descrito a la evaluación formativa como un conjunto de estrategias pedagógicas que permite a docentes y a estudiantes describir, analizar, valorar y orientar la progresión de los aprendizajes estudiantiles según objetivos previamente determinados (Lipnevich et al., 2016; Panadero et al., 2012; Shavelson et al., 2008). De esta forma, la retroalimentación se constituye en un puente entre el aprendizaje y la evaluación formativa, como también entre la enseñanza y la evaluación formativa, en tanto nutre de información sobre su desempeño al profesorado y al estudiantado (Bond et al., 2020).

### Estrategias para la evaluación formativa: retroalimentación docente, coevaluación y autoevaluación

Investigaciones sobre la calidad de la enseñanza han indicado que la evaluación formativa posee un alto impacto en el desarrollo de los aprendizajes estudiantiles (Klute et al., 2017) ya sea en educación primaria, secundaria o superior (López-Pastor y Pérez-Pueyo, 2017). La literatura ha destacado varias estrategias de evaluación formativa (Moos y Brookhart, 2019). En el presente trabajo se ha optado por recoger tres estrategias que se han indicado en la literatura como representativas de la evaluación formativa: retroalimentación docente, coevaluación y autoevaluación (Bond et al., 2020; Popham, 2013). Por ahora, se han dejado fuera a otras estrategias, como, por ejemplo, aquellas que generan situaciones que evidencian el aprendizaje. La retroalimentación docente se ha definido como la valoración de un profesor respecto del desempeño de un estudiante en una tarea. Se ha demostrado que su principal contribución radica en que junto con informar al estudiante sobre qué acciones debe desarrollar de mejor modo a partir de identificar errores, también

orienta respecto de cómo hacer para obtener un mejor desempeño (Panadero y Lipnevich, 2022). La coevaluación refiere a la valoración del desempeño de un estudiante que se realiza entre compañeros y compañeras de clase (Panadero et al., 2023). Y la autoevaluación señala la valoración que cada estudiante realiza respecto de su propio desempeño (Harris y Brown, 2022). Brown y Harris (2013) y Sánchez et. al (2017) han reportado que el estudiantado que ha participado de procesos de autoevaluación y coevaluación ha obtenido mejores resultados de aprendizaje en pruebas posteriores que aquellos que no lo hicieron, demostrando la utilidad de ambas prácticas. De las tres estrategias descritas la retroalimentación docente posee un rol central ya que surge desde un docente experto, orienta la práctica del estudiante hacia la mejora y permite atender a las diferencias individuales del estudiantado (Andrade, 2023; Hooley y Thorpe, 2017).

### **Calidad de la retroalimentación para el aprendizaje**

El interés académico por la comprensión de la retroalimentación docente se ha incrementado y diversos estudios han analizado qué factores se asocian a una retroalimentación de calidad. Por ejemplo, Ossenberg et al., (2019) han analizado 61 publicaciones y han identificado diez atributos de una retroalimentación que mejoran el desempeño del estudiantado, por ejemplo: que es detallada y que considera las necesidades estudiantiles. Adarkwah (2021) ha realizado una revisión de alcance donde señala que la literatura clasifica a la retroalimentación en dos tipos: retroalimentación formativa, que describe el progreso del desempeño estudiantil en términos cualitativos y una retroalimentación sumativa, que clasifica el desempeño en función de una cuantificación del dominio mediante calificaciones, habitualmente con una finalidad acreditativa. Ha concluido que una retroalimentación de calidad es aquella que describe con precisión en qué aspectos del desempeño mejorar y que indica con claridad qué hacer para avanzar, más allá de la cuantificación de aprendizajes. Por su parte, Tay y Lam (2022) han estudiado el impacto de distintas estrategias de retroalimentación en 75 estudiantes de secundaria en Singapur. Han concluido que una retroalimentación de calidad promueve el compromiso del estudiantado con la retroalimentación recibida si en ella se visualizan aprendizajes que ocurrirán en el futuro y se considera la implicancia emocional que aquello tendrá. Finalmente, Panadero y Lipnevich (2022) han analizado 14 modelos de retroalimentación en una revisión sistemática. Desde dicho trabajo han elaborado el modelo MISCA (Sigla del idioma inglés) compuesto por cinco factores: Mensaje, Implementación, Estudiante, Contexto y Agentes. Este modelo enfatiza que una retroalimentación de calidad tiene en cuenta las características del estudiantado, quienes junto con recibir retroalimentación también son productores de retroalimentación para sí mismos en la autoevaluación y productores de retroalimentación para sus compañeros y

compañeras de clase en la coevaluación. Respecto del contexto, se ha indicado que una retroalimentación de calidad requiere que cada centro educativo propicie climas escolares positivos y climas de aula orientados al aprendizaje (Heritage, 2010) ya que una retroalimentación efectiva incide favorablemente en el desempeño estudiantil si consigue afectar positivamente en sus estados motivacionales (Rowe, 2017). Del mismo modo se ha señalado que los sistemas educativos nacionales pueden favorecer el desarrollo de la evaluación formativa, y dentro de ella a la retroalimentación, o bien obstaculizar su implementación (Bond et al., 2020; van der Kleij et al., 2018; van der Kleij y DeLuca, 2023).

## Justificación de la investigación

Aunque los trabajos descritos ofrecen un valioso aporte a la comprensión de la evaluación formativa y de la calidad de la retroalimentación, la cantidad de estudios sobre la temática tiende a ser mayor en educación superior que en primaria y en secundaria (Alqassab et al., 2023; Sánchez et al., 2017). Por otro lado, la operativización de los procesos de retroalimentación aún es confusa y requiere de un ordenamiento conceptual mayor (Hortigüela et al., 2019; Van der Kleij et al., 2018). Probablemente, lo anterior se relaciona con que en algunos contextos educativos la retroalimentación se utiliza con varios objetivos, ya sean formativos, sumativos (calificar con notas) o mixtos; lo que de algún modo constituye en un obstáculo para monitorear el uso de esta estrategia y para conocer su impacto en los aprendizajes (Adarkwah, 2021). En el presente trabajo se busca responder a los desafíos descritos. El principal objetivo de esta investigación es diseñar dos escalas de autorreporte docente con una estructura teórica clara y validez empírica. Una escala para evaluar la realización de Estrategias de Evaluación Formativa y otra escala para evaluar la Calidad de la Retroalimentación para el Aprendizaje que ofrece el profesorado al estudiantado en las clases. El desarrollo de ambas escalas puede ser de utilidad para que los centros educativos inicien o desarrollen procesos de comprensión de los procesos de evaluación formativa y retroalimentación de los aprendizajes o para el desarrollo de programas de asesoramiento en evaluación formativa destinada a docentes en formación y en ejercicio (Matthews et al., 2023; Pat-El et al., 2013; Shavelson et al., 2008).

## MÉTODO

### Participantes

El grupo de participantes se compone de 364 docentes que ejercen docencia en la Región de Tarapacá, Chile. Los docentes enseñan diversas materias en cursos desde 1° hasta 12° grado. El promedio de edad es de 38 años. Han participado

242 mujeres (66.5 %) y 122 hombres (33.5 %). Pertenecen a un total de 13 escuelas, 9 escuelas privadas con financiamiento público y 4 escuelas públicas. En las escuelas privadas con financiamiento público se desempeñan 278 docentes (76.4%) y 86 trabajan en escuelas públicas (23.6 %). En el año 2021, el total de docentes escolares de la Región de Tarapacá ha correspondido a 4607, el 1.8 % del total de docentes escolares de Chile, y atienden aproximadamente a 85200 estudiantes (Centro de estudios MINEDUC, 2022). Mayoritariamente, los docentes participantes pertenecen a escuelas que se orientan hacia la innovación docente y la mejora educativa, un aspecto reconocido formalmente por las políticas públicas educativas chilenas. Aunque aquello puede constituir una limitación, se ha considerado oportuno validar estas escalas con docentes orientados a la mejora educativa para explorar adecuadamente la validez de constructos complejos que están relacionados con dicha mejora.

## Diseño de los instrumentos

Para nutrir de contenido teórico el diseño de la Escala de Estrategias de Evaluación Formativa (E3F) se ha considerado la conceptualización de Popham (2013) y Shavelson et al. (2008). Dicho marco ha orientado la construcción de ítems distribuidos en tres subescalas: Retroalimentación docente, Coevaluación y Autoevaluación. Los autores del presente trabajo, teniendo en cuenta este marco teórico, han elaborado un conjunto de ítems que describen acciones propias de dichas estrategias, que consideren procedimientos de evaluación como el uso de rúbricas, pautas o tutorías (Andrade, 2023) y que enfatizen el papel de la reflexión personal o que mostrasen la versión opuesta de sistemas tradicionales de evaluación. El proceso de validación empírica posterior ha sido útil para seleccionar los ítems que finalmente han funcionado según estos criterios.

Para el diseño de la Escala de Calidad de la Retroalimentación para el Aprendizaje (ECRA) se han seleccionado cuatro revisiones de la literatura, según los siguientes criterios: son sistematizaciones teóricas o revisiones sistemáticas, fueron publicadas en los últimos cinco años y consideran el impacto de los procesos de retroalimentación docente en el estudiantado. Estos artículos han sido descritos en el apartado introductorio. En el Cuadro 1 se puede apreciar que se ha realizado una síntesis a partir de los contenidos de estos cuatro trabajos en dimensiones que son coherentes con el modo con que la literatura actual señala los objetivos de la retroalimentación: el desempeño, la motivación y la autorregulación del aprendizaje (Lipnevich y Panadero, 2021). En dicha síntesis es posible observar la configuración de tres dimensiones: una dimensión instruccional que enfatiza en la transmisión del mensaje y en el contenido de la retroalimentación, una dimensión interaccional que destaca el valor de lo motivacional y de empatizar con las acciones y afectos

del estudiantado y una dimensión autorregulatoria orientada a promover que el estudiantado reflexione sobre su propio involucramiento y su desempeño en las tareas de aprendizaje.

En su conjunto, ambos marcos conceptuales ofrecen una estructura teórica sólida y sustentan el diseño de una base de ítems para cada escala. En el caso de la Escala de Estrategias de Evaluación Formativa se construyó una base inicial de 12 ítems, cuatro ítems para tres subescalas. Respecto de la Escala de Calidad de la Retroalimentación para el Aprendizaje se elaboró una base inicial de 22 ítems distribuidos en tres subescalas, siete ítems representativos de la dimensión instruccional, siete ítems representativos de la dimensión interaccional y ocho ítems representativos de la dimensión autorregulatoria. Ambas escalas contienen ítems directos e inversos, poseen un formato de respuesta tipo Likert con rangos de 1 a 5, y expresan estar *totalmente en desacuerdo* hasta *totalmente de acuerdo*.

## Procedimiento

Inicialmente, se ha contactado a los equipos directivos de cada escuela y se les ha propuesto participar en el proyecto de investigación. El profesorado ha participado voluntariamente y ha aprobado un consentimiento informado que daba cuenta de la anonimización y confidencialidad de los datos. Estos procedimientos han sido aprobados por el Comité de Ética de la Investigación de la Universidad Autónoma de Madrid, informe CEI-125- 2566. La recolección de datos ha sido la misma en cada uno de los centros educativos. El profesorado se ha reunido en un lugar habilitado con computadores y han contestado individualmente las escalas en una plataforma virtual.

## Análisis de datos y procedimientos de validación

En primer lugar, se han realizado análisis factoriales exploratorios (AFE) que han orientado la configuración de las subescalas de ambas escalas. Los AFE se han realizado en submuestras aleatorias correspondientes a la mitad del tamaño muestral total ( $N = 182$ ). Para ello, se ha utilizado el método de componentes principales con rotación Equamax. Para la identificación de factores se han considerado valores de carga factorial iguales o superiores a .40. Posteriormente, se han obtenido los indicadores de la prueba KMO, de la prueba de esfericidad de Bartlett y de la varianza total explicada de cada escala y sus subescalas (Lloret-Segura, 2014). Aquello ha permitido la selección de ítems y la configuración factorial para ambas escalas.

En segundo lugar, en las restantes submuestras aleatorias de la mitad del tamaño muestral total ( $N = 182$ ), los ítems seleccionados para ambas escalas han sido analizados mediante un análisis factorial confirmatorio (AFC) con modelos de ecuaciones estructurales (SEM) a partir de diseños originales, utilizando el método de máxima verosimilitud (ML) y siguiendo las consideraciones de Ruiz et. al (2020). Luego, en cada escala, se ha obtenido la media de cada ítem, su desviación típica, el error típico y el estadístico de la prueba  $t$  para una muestra que ha comparado la media obtenida con el valor central del rango de respuestas.

A continuación, para el conjunto de datos de cada escala, se ha obtenido el índice de normalidad multivariante de Mardia, mediante los indicadores de asimetría y curtosis, su valor y su rango críticos. Una vez definidos los modelos definitivos se han obtenido los estadísticos alfa de Cronbach, omega de McDonald. Consecuentemente, se ha determinado la bondad de ajuste y la varianza explicada de cada instrumento. Después se ha realizado un análisis de validación cruzada para cada escala entre dos submuestras aleatorias de la mitad del tamaño muestral total. Finalmente, se han analizado las diferencias según el nivel educativo donde se desempeñan los docentes participantes, primaria o secundaria, y diferencias según género. Para el análisis de datos se ha utilizado el paquete estadístico IBM SPSS y Amos versión 28.

## RESULTADOS

### Análisis factoriales exploratorios

Se ha realizado un AFE con los 12 ítems de la Escala Estrategias de Evaluación Formativa, con el objetivo de identificar la contribución de cada ítem a su configuración unifactorial. Dicho análisis ha presentado una varianza total explicada del 27.9 %, un valor KMO = .76 y un valor  $p$  de la prueba de Bartlett =  $p < 0.01$ . Allí se ha señalado que dos ítems no han alcanzado el peso factorial de .40, E3F\_06 = .33 y E3F\_12 = .38 y se han eliminado del conjunto. Posteriormente, se ha realizado un nuevo AFE, sin estos dos ítems y ha mostrado una varianza total explicada de 31.6 %, un valor KMO = .76 y un valor  $p$  de la prueba de Bartlett =  $p < 0.01$ . Luego, en una submuestra aleatoria compuesta por la mitad de los casos del tamaño muestral total ( $N = 182$ ), con los 10 ítems seleccionados, se ha realizado un nuevo AFE considerando la configuración de tres factores, una varianza total explicada del 57.3 %, un valor KMO = .71 y un valor  $p$  de la prueba de Bartlett =  $p < 0.01$ . En este AFE, se ha eliminado el ítem E3F\_01 ya que es el ítem inverso del ítem E3F\_07 y ambos figuran en el mismo factor. De este modo, con los 9 ítems seleccionados se ha realizado un nuevo AFE en las mismas condiciones que el anterior. Este AFE ha mostrado una varianza total explicada del 59.1 %, un valor KMO = .69 y un valor  $p$  de la prueba de Bartlett =  $p < 0.01$ . El Factor 1 contiene afirmaciones que



refieren a la retroalimentación docente y a la autoevaluación. El Factor 2 posee ítems que refieren a la coevaluación. Y el Factor 3 refiere fundamentalmente al uso de rúbricas. El resultado fue un factor que explicaba el 31 % de la varianza que explicaba retroalimentación docente y autoevaluación y un factor que explicaba el 46.9 % de la varianza que ha referido a coevaluación. Para el AFC, junto a los pesos factoriales, se ha considerado la adscripción del ítem según el modelo teórico que se usó para construir el cuestionario. Así se ha actuado para decidir la adscripción en los ítems E3F\_04, E3F\_07 y ECRA\_09R y E3F\_05R.

**Tabla 1**

*Análisis factorial exploratorio de la Escala de Estrategias de Evaluación Formativa*

Ítem	Factor 1	Factor 2	Factor 3
E3F_04: En cada evaluación utilizo algún procedimiento -rúbrica, pauta, tutoría- que informa al estudiante el nivel de dominio que alcanzó en la realización de una tarea.			.60
E3F_03R: No acostumbro a informar a mis estudiantes sobre qué pueden hacer para aprender mejor.	.77		
E3F_10R: Me resulta muy complicado informar a cada uno de mis estudiantes sobre qué errores cometieron en las evaluaciones.	.65		
E3F_08R: Evito que mis estudiantes comenten el trabajo de sus compañeros y compañeras durante las clases.		.83	
E3F_07: Solicito a mis estudiantes que evalúen el trabajo de sus compañeros y compañeras mediante rúbricas o pautas.		.52	.59
E3F_11R: Evito pedir a mis estudiantes que analicen el desempeño de sus compañeros y compañeras.		.84	
E3F_09R: Prefiero evitar que mis estudiantes se evalúen a sí mismos o a sí mismas.	.63		.39
E3F_02: Pido a mis estudiantes que evalúen su propio desempeño mediante rúbricas o pautas.			.73
E3F_05R: No utilizo la autoevaluación porque mis estudiantes se califican mejor de lo que realmente corresponde.	.54		.47
Varianza explicada	31.2 %	15.6 %	12.2 %

Nota. N = 182.

**Cuadro 1**  
*Síntesis de variables de la calidad de la retroalimentación docente para el aprendizaje a partir de la comparación de investigaciones previas*

Revisiones de la literatura sobre retroalimentación del profesorado al estudiantado			Dimensiones
Ossenberg et. al (2019)	Adarkwah (2021)	Tay y Lam (2022)	Panadero y Lipnevich (2022)
Forma parte de un proceso	Es detallada	Ofrece comentarios específicos y generales sobre el desempeño en una tarea	Mensaje: la información contenida en el mensaje de retroalimentación debe ser clara y útil
Se basa en criterios	Es clara y simple		Contexto: el enfoque pedagógico y el modo de entrega de la retroalimentación favorecen el involucramiento
Utiliza múltiples fuentes de evidencia			
Es frecuente		Orienta y motiva la práctica hacia la mejora	
Implica una interacción hábil	Orienta la acción del estudiante hacia la mejora		
Involucra al estudiante en más de una forma	Utiliza un tono de voz adecuado	Presta atención a la respuesta emocional del estudiantado	Estudiante: considera las diferencias individuales, creencias motivacionales, conocimientos previos, género, diferencias culturales, autoeficacia, etc.
Se adapta a las necesidades del estudiantado	Es adecuada a las demandas del estudiantado		Dimensión interaccional Orientada a la motivación
Es bienvenida por el estudiantado			
Es bidireccional o recíproca			
Se centra en el futuro		Presta atención a futuros procesos de evaluación de aprendizajes	Dimensión autorregulatoria Orientada a la autorregulación
Es oportuna	Es oportuna	Es oportuna	
			Agentes: promueve la integración del docente, los pares y el o la aprendiz

*Nota.* Elaboración propia a partir de los trabajos que se indican.

Respecto de la Escala de Calidad de la Retroalimentación, también se ha realizado un AFE. Inicialmente, en una submuestra aleatoria con la mitad de los casos ( $N = 182$ ), se ha realizado un AFE con los ítems de cada subescala por separado, a partir de la síntesis del Cuadro 1. Para ello, se han considerado aceptables valores de carga factorial superiores a .50. De los 7 ítems que correspondían a la dimensión instruccional, tres de ellos no han alcanzado el peso factorial señalado y han sido eliminados de este conjunto (ECRA\_09R = .28, ECRA\_19R = .46 y ECRA\_20R = .46). El AFE de esta subescala presenta un valor KMO = .66, un valor  $p$  de Bartlet =  $p < 0.01$  y una varianza total explicada del 32.5 %. De los 7 ítems que correspondían a la dimensión interaccional, dos de ellos no han alcanzado el peso factorial señalado y han sido eliminados (ECRA\_10R = .13 y ECRA\_15R = .47). El AFE de esta subescala presenta un valor KMO = .65, un valor  $p$  de Bartlet =  $p < 0.01$ . y una varianza total explicada del 31.4 %. Finalmente, de los 8 ítems a la dimensión autorregulatoria, dos de ellos no han alcanzado el peso factorial señalado y han sido eliminados (ECRA\_12 = .42, ECRA\_21 = .48). El AFE de esta subescala presenta un valor KMO = .76, un valor  $p$  de Bartlet =  $p < 0.01$ . y una varianza total explicada del 37.5. Con un total de 15 ítems se ha realizado un nuevo AFC un valor KMO = .81, un valor  $p$  de Bartlet =  $p < 0.01$ , y una varianza total explicada del 48.7 %. En este análisis los ítems ECRA\_22R, ECRA\_13 y ECRA\_2R han sido eliminados por tratarse de ítems inversos de los ítems ECRA\_16, ECRA\_17 Y ECRA\_3, correspondientemente. Las cargas factoriales de los ítems en los tres factores del AFE de la muestra total son muy semejantes a los que se muestran en esta submuestra.

**Tabla 2**

*Análisis factorial exploratorio de la Escala Calidad de la Retroalimentación para el Aprendizaje*

Ítem	Factor 1	Factor 2	Factor 3
ECRA_01: Muestro a los estudiantes cómo se hacen o ejecutan las tareas de aprendizaje que les propongo.	.59		
ECRA_07: Mis comentarios u observaciones son claros, fáciles de entender y contienen instrucciones sobre qué hacer para lograr los objetivos de aprendizaje.	.70		
ECRA_08: Cuando informo a un estudiante sobre el resultado de su tarea uso un tono fraterno, respetuoso y cordial.	.45	.49	
ECRA_18: Mis comentarios sobre el desempeño de los estudiantes son específicos y abordan acciones importantes sobre las que pueden que mejorar.	.66		

Ítem	Factor 1	Factor 2	Factor 3
ECRA_05: Considero las características individuales de mis estudiantes (contexto social, características de personalidad, necesidades especiales) para comentar sus trabajos en función de sus particularidades.	.46	.36	
ECRA_04: Casi siempre me pongo en el lugar de mis estudiantes y trato de pensar en cómo afrontan las tareas de aprendizaje que les propongo.	.40	.58	
ECRA_03: Estoy dispuesto a recibir comentarios de mis estudiantes sobre el resultado de sus evaluaciones.		.79	
ECRA_06R: Evito pensar en cómo se sienten mis estudiantes en las tareas de aprendizaje que les propongo.		.56	.51
ECRA_11R: Destino poco esfuerzo a preguntar las expectativas que los estudiantes tienen sobre las tareas de aprendizaje que les propongo.			.82
ECRA_14R: Casi nunca pregunto al estudiante sobre lo piensa de su propio desempeño en la ejecución de las tareas.			.75
ECRA_16: No dejo pasar mucho tiempo para entregar a mis estudiantes las observaciones o comentarios sobre sus trabajos.	.32		.31
ECRA_17: Indico a los estudiantes qué procedimientos o tareas realizaron correctamente para que reconozcan sus logros.	.54		.44
Varianza explicada	33.2%	10.4%	9.0%

Nota.  $N = 182$ .

Con 12 ítems, se ha realizado un nuevo AFE destinado a conocer la trifactorialidad. Se ha utilizado el método de componentes principales y rotación Equamax. Para el AFC, junto a los pesos factoriales, se ha considerado la adscripción del ítem según el modelo teórico que se usó para construir el cuestionario. Así se ha actuado para decidir la adscripción especialmente en los ítems ECRA\_6, ECRA\_16 y ECRA\_17. Esta estructura se confirma con el AFE de la muestra total. Dicho análisis presenta un valor KMO = .83, un valor  $p$  de Bartlett =  $p < 0.01$  y una varianza total explicada del 52.7 %. En la Tabla 2 se observan los ítems del Factor 1, de la dimensión instruccional y que conforman la subescala de instrucción focalizada, los ítems del Factor 2, que corresponden a la dimensión interaccional y que configuran la subescala interacción empática, y los ítems del Factor 3, que provienen de la

dimensión autorregulatoria y que conforman la subescala autorregulación del desempeño estudiantil.

### **Análisis factoriales confirmatorios: bondad de ajuste, varianza explicada y análisis descriptivos**

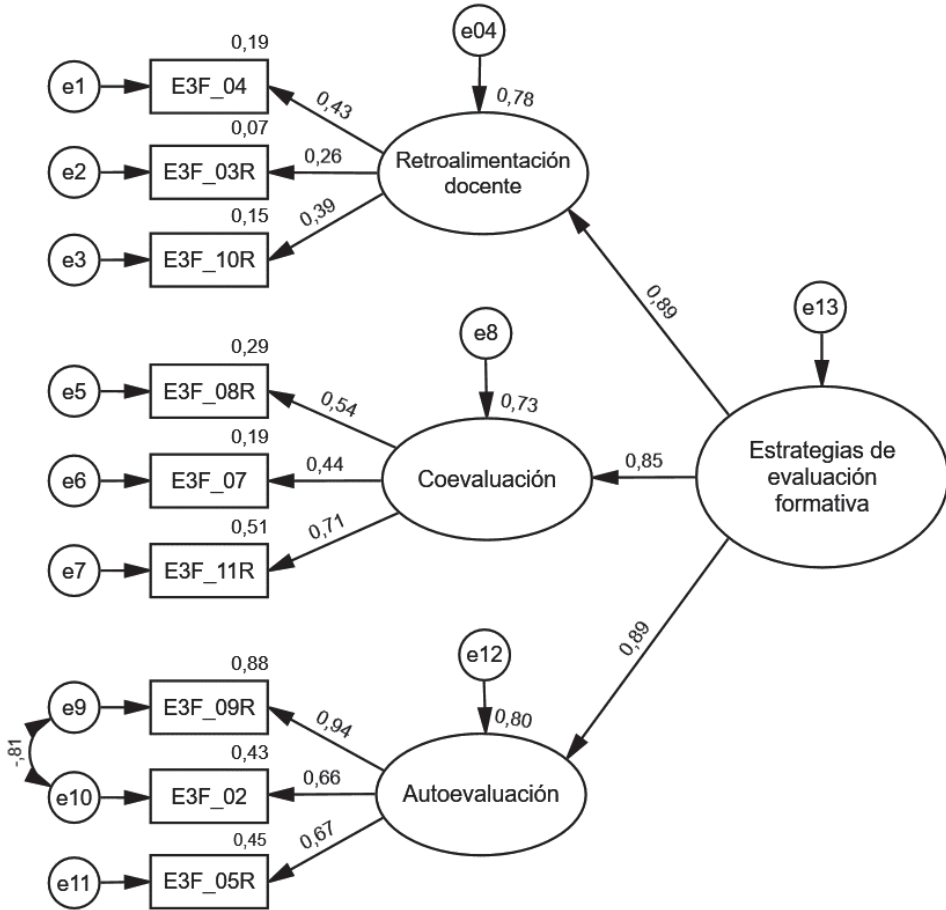
A partir del AFE descrito en la Tabla 1, se ha realizado un AFC con los 9 ítems seleccionados de la Escala de Estrategias de Evaluación Formativa en una submuestra aleatoria con la otra mitad de los casos de la muestra total ( $N = 182$ ). Se obtuvo un modelo original que presenta indicadores cercanos a los siguientes criterios de referencia para la bondad de ajuste de modelos SEM:  $\chi^2 / gl < 3$ ,  $CFI \geq .95$  y  $RMSEA \leq .05$ . Los indicadores de bondad de ajuste que presenta el modelo SEM de dicha escala son aceptables:  $\chi^2 / gl = 1.3$ ,  $CFI = .97$  y  $RMSEA = .04$ ;  $IC (.05 - .11)$ . El modelo estructural resultante se presenta en la Figura 1 y contiene tres variables latentes que representan a las subescalas: Retroalimentación docente, compuesta por tres ítems; Coevaluación, compuesta por tres ítems, Autoevaluación, compuesta por tres ítems y una variable latente de segundo orden.

Respecto de la Escala Calidad de la Retroalimentación para el Aprendizaje, a partir del AFE previo, también se ha realizado un AFC en una submuestra aleatoria con esta mitad de los casos de la muestra total ( $N = 182$ ). Este análisis ha considerado a los 12 ítems que se presentan en la Tabla 2. De este modo, se obtuvo un modelo original que responde satisfactoriamente a los indicadores de bondad de ajuste descritos en el párrafo precedente:  $\chi^2 / gl = 1.2$ ,  $CFI = .96$  y  $RMSEA = .03$ ;  $IC (.00 - .08)$ . En la Figura 2 se presenta el modelo estructural resultante. Contiene tres variables latentes que representan a las subescalas: Instrucción focalizada, que emerge de la dimensión interaccional y que contiene cinco ítems, Interacción empática que surge de la dimensión interaccional y que contiene tres ítems y Autorregulación del desempeño estudiantil que surge de la dimensión autorregulatoria y que contiene cuatro ítems; y una variable latente de segundo orden.

La Figura 1 presenta los pesos estandarizados y los coeficientes de determinación o  $R^2$  de cada ítem y de cada variable latente de las tres subescalas de la Escala de Estrategias de Evaluación Formativa. La mayoría de los ítems supera el indicador de peso estandarizado .40, salvo los ítems E3F\_03R y E3F\_10R. Los valores de  $R^2$  en las subescalas es alto, siendo menor en el caso de la estrategia Coevaluación con un indicador de .73. Las medias y valores de la prueba  $t$  presentes en la Tabla 3 señalan valores que se ubican significativamente por sobre el valor central del rango de respuestas.

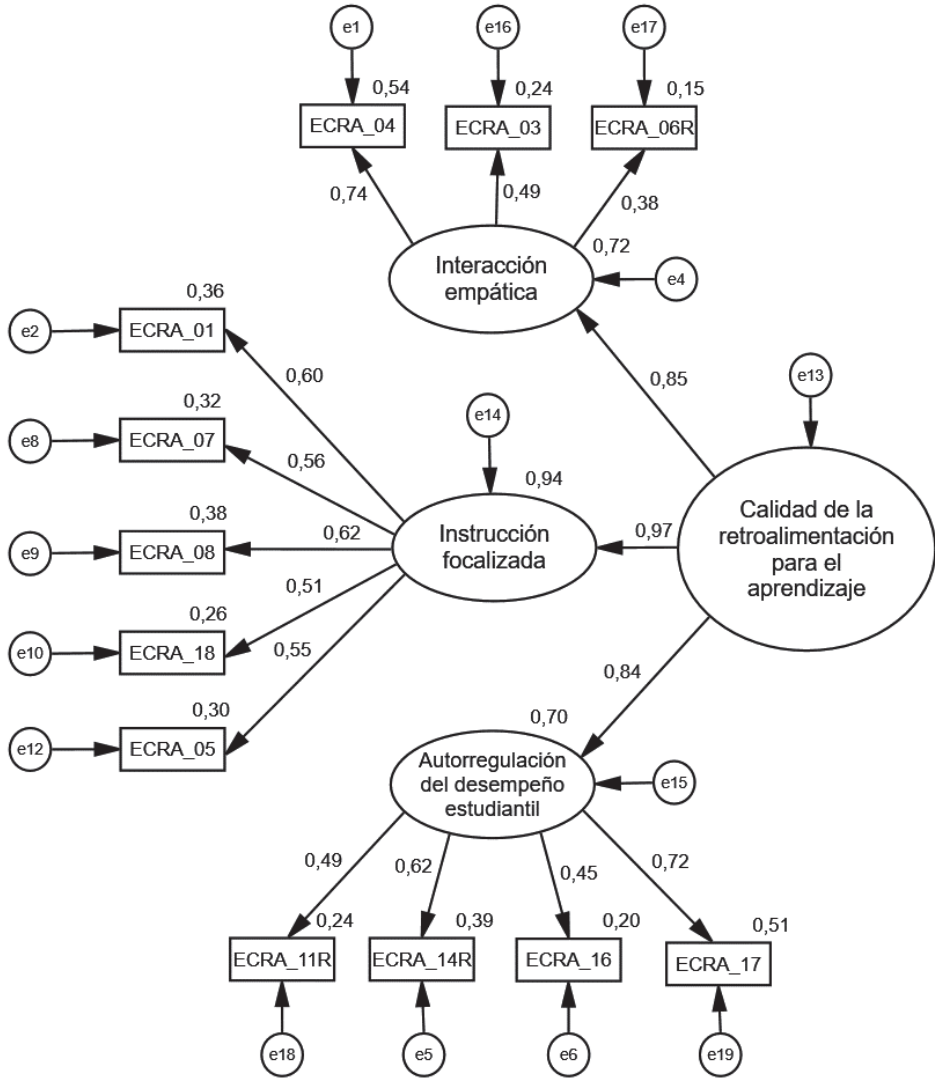
**Figura 1**

Modelo estructural de la Escala de Estrategias de Evaluación Formativa



**Figura 2**

Modelo estructural de la Escala Calidad de la Retroalimentación para el Aprendizaje



**Tabla 3**

*Medias, desviaciones estándar, error típico y estadísticos de la prueba t de Student para una muestra de la Escala de Estrategias de Evaluación Formativa*

Ítem	M	DE	ET	t	gl	p
E3F_04	4.3	.76	0.05	23.8	181	<.001
E3F_03R	4.4	.83	0.06	22.4	181	<.001
E3F_10R	3.8	1.0	0.08	10.9	181	<.001
E3F_08R	3.4	1.1	0.08	5.12	181	<.001
E3F_07	3.1	1.2	0.09	2.02	181	.045
E3F_11R	3.4	1.1	0.08	5.57	181	<.001
E3F_09R	3.9	1.0	0.07	11.7	181	<.001
E3F_02	3.9	1.0	0.07	12.0	181	<.001
E3F_05R	3.8	1.0	0.08	10.6	181	<.001
E3F_RT	4.2	.61	0.04	26.4	181	<.001
E3F_CV	3.6	.91	0.06	9.02	181	<.001
E3F_OFF	3.8	.93	0.06	12.8	181	<.001
E3F Total	3.9	.58	0.04	20.9	181	<.001

*Nota.* N = 182. M = Media; DE= Desviación Estándar; ET= Error Típico; t = Valor t; p = Prueba t de Student para una muestra con valor de prueba = 3 y nivel de significación .05; E3F\_RT = Subescala retroalimentación docente; E3F\_CV = Subescala coevaluación; E3F\_AV = Subescala autoevaluación; E3F = Escala Estrategias de Evaluación Formativa.

**Tabla 4**

*Medias, desviaciones estándar, error típico y estadísticos de la prueba t de Student para una muestra de la Escala Calidad de la Retroalimentación para el Aprendizaje*

Ítem	M	DE	ET	t	gl	p
ECRA_01	4.7	.49	.03	60.0	181	<.001
ECRA_07	4.3	.73	.03	37.8	181	<.001
ECRA_08	4.7	.55	.03	65.3	181	<.001
ECRA_18	4.2	.64	.04	31.7	181	<.001
ECRA_05	4.5	.97	.03	43.8	181	<.001
ECRA_04	4.5	1.0	.03	52.6	181	<.001
ECRA_03	4.6	.64	.03	49.1	181	<.001



Ítem	<i>M</i>	<i>DE</i>	<i>ET</i>	<i>t</i>	<i>gl</i>	<i>p</i>
ECRA_06R	4.3	1.0	.05	24.9	181	<.001
ECRA_11R	3.9	.96	.05	16.1	181	<.001
ECRA_14R	4.1	.62	.05	20.8	181	<.001
ECRA_16	4.0	.42	.05	19.1	181	<.001
ECRA_17	4.5	.47	.03	46.0	181	<.001
ECRA_IF	4.5	.64	.02	67.4	181	<.001
ECRA_ES	4.5	.41	.02	60.6	181	<.001
ECRA_ADE	4.1	.49	.03	32.6	181	<.001
Total ECRA	4.3	.73	.02	62.5	181	<.001

*Nota.* *N* = 182. *M* = Media; *DE*= Desviación estándar; *ET*= Error típico; *t* = Valor *t*; *p* = Prueba *t* de Student para una muestra con valor de prueba = 3 y nivel de significación .05; ECRA\_OF = Subescala instrucción focalizada; ECRA\_IE = Subescala interacción empática; ECRA\_ADE = Subescala autorregulación del desempeño estudiantil; ECRA = Escala Calidad de la Retroalimentación para el Aprendizaje.

La Figura 2 presenta los pesos estandarizados y los coeficientes de determinación o  $R^2$  de cada ítem y de cada variable latente de las tres subescalas de la Escala de Calidad de la Retroalimentación para el Aprendizaje. Se observa que la mayoría de los ítems supera el indicador de peso de medida estandarizado de .40, salvo el ítem ECRA\_06 = .38. Los valores de  $R^2$  en las subescalas es alto, siendo menor en el caso de la subescala autorregulación del desempeño estudiantil con un indicador de .70. Las medias y los valores de la prueba *t* de Student para una muestra que se presentan en la Tabla 4 señalan que las puntuaciones se ubican significativamente por sobre el valor central del rango de respuestas.

### Análisis de normalidad

Los indicadores de normalidad multivariante de Mardía (1974) evidencian que la distribución de la E3F es cercana a la normalidad y que la ECRA no es normal. En el caso de la asimetría, el valor del estadístico es mayor que su valor crítico. Respecto de la curtosis, el valor del estadístico no se ubica dentro del rango crítico establecido por Mardía según el tamaño muestral (Wulandari et al., 2021). Ya que SEM con ML solicita normalidad se obtuvo el valor *p* del índice de Bollen-Stine (BS) con dos mil bootstrap (Cheung y Lau, 2008; Fan, 2003). Se ha obtenido un valor *p* del índice BS de .06. Los indicadores *p* del Índice de BS corrigen la anormalidad detectada al superar el valor *p* de .05.

## Análisis de confiabilidad

Los valores Alfa de Cronbach señalan una confiabilidad por consistencia interna aceptable u óptimas ya que son iguales o superiores a .70. La Escala de Estrategias de Evaluación Formativa ha obtenido un alfa de Cronbach de .71 y la Escala de Calidad de la Retroalimentación para el Aprendizaje ha obtenido alfa de Cronbach de .78. Los valores del indicador omega de McDonald son idénticos a los de alfa de Cronbach en ambas escalas. Las correlaciones entre escalas y subescalas muestran la ausencia de colinealidad y una relación significativa entre las subescalas de cada escala (Tabla 5).

**Tabla 5**

*Estadísticos de correlación múltiple entre las escalas y subescalas*

Variable	E3F	E3F_RT	E3F_CV	E3F_AV	ECRA	ECRA_IF	ECRA_ES
E3F_RT	.69*						
E3F_CV	.78*	.25*					
E3F_AV	.74*	.38*	.36*				
ECRA	.42*	.51*	.19*	.29*			
ECRA_IF	.31*	.44*	.10*	.20*	.79*		
ECRA_IE	.31*	.30*	.17*	.25*	.70*	.38*	
ECRA_ADE	.37*	.45*	.18*	.24*	.86*	.50*	.44*

*Nota.*  $N = 364$ . \*La correlación es significativa en el nivel .01 bilateral; E3F = Escala de Estrategias de Evaluación Formativa; E3F\_RT = Subescala retroalimentación docente; E3F\_CV = Subescala coevaluación; E3F\_AV = Subescala autoevaluación; ECRA = Escala calidad de la retroalimentación para el aprendizaje; Escala ECRA\_IF = Subescala instrucción focalizada; ECRA\_IE = Subescala interacción empática; ECRA\_AD = Subescala autorregulación del desempeño estudiantil.

## Validación cruzada

Al segmentar la muestra total en dos submuestras aleatorias de igual tamaño se obtienen indicadores de bondad de ajuste cercanos a los criterios satisfactorios en ambas escalas (Tabla 6).

**Tabla 6**

*Análisis factoriales confirmatorios con submuestras aleatorias*

E3F					ECRA				
Modelo	$\chi^2$ /gl	CFI	RMSEA	IC	Modelo	$\chi^2$ /gl	CFI	RMSEA	IC
E3F-1 (N = 182)	2.5	.91	.08	(.05 - .11)	ECRA-1 (N = 182)	1.2	.96	.03	(.00 -.06)
E3F-2 (N = 182)	2.2	.90	.08	(.05 - .01)	ECRA-2 (N = 182)	1.8	.90	.06	(.04 -.09)
E3F (N = 364)	3.2	.91	.07	(.05 - .09)	ECRA (N = 364)	1.8	.94	.05	(.03 -.06)

*Nota.* E3F = Escala de Estrategias de Evaluación Formativa; ECRA = Escala Calidad de la Retroalimentación para el Aprendizaje.

### Análisis de diferencias según nivel educativo

A partir de análisis de la prueba ANOVA se ha explorado la existencia de diferencias significativas entre las medias de ambas escalas y los tres niveles educativos donde se trabaja el profesorado participante, primaria 1° a 4°, primaria 5° a 8° y secundaria 9° a 12°. Los resultados señalan que no existe diferencia significativa según nivel educativo, en ambas escalas (Tablas 7 y 8).

**Tabla 7**

*Estadísticos de la prueba ANOVA para la comparación de medias según el nivel educativo*

Escala	Nivel	M	DE	N	F	p
Escala de Estrategias de Evaluación Formativa	Primaria 1° a 4°	3.8	.60	81	.97	.37
	Primaria 5° a 8°	3.9	.59	110		
	Secundaria 9° a 12°	3.9	.59	173		
	Total	3.8	.59	364		

**Tabla 8**

*Estadísticos de la prueba ANOVA para la comparación de medias según nivel educativo*

Escala	Nivel	M	DE	N	F	p
Escala Calidad de la Retroalimentación para el Aprendizaje	Primaria 1° a 4°	4.3	.41	81	.124	.94
	Primaria 5° a 8°	4.3	.42	110		
	Secundaria 9° a 12°	4.3	.40	173		
	Total	4.3	.41	364		

## Análisis de diferencias según género

Para comprobar la incidencia del género del profesorado en las medias de ambas escalas se han realizado análisis con la prueba t de Student para muestras independientes. Los resultados han señalado que en ambas escalas la diferencia de medias entre profesoras y profesores no es significativa (Tablas 9 y 10).

**Tabla 9**

*Estadísticos de la prueba t de Student para la diferencias de medias según género*

Escala	Género	M	DE	N	F	p
Escala de Estrategias de Evaluación Formativa	Mujeres	3.9	.58	242	.27	.60
	Hombres	3.7	.61	122		
	Total	3.8	.59	364		

**Tabla 10**

*Estadísticos de la prueba t de Student para la diferencia de medias según género*

Subescala	N	M	DE	N	F	p
Escala Calidad de la Retroalimentación para el Aprendizaje	Mujeres	4.3	.38	242	1.7	.18
	Hombres	4.2	.44	122		
	Total	4.3	.41	364		

## DISCUSIÓN

Las puntuaciones medias de ambas escalas indican que el grupo de docentes señala realizar una evaluación formativa y una retroalimentación de calidad aceptable. Un resultado previsible, considerando que los docentes participantes trabajan en escuelas que se orientan a la innovación docente. Ambos instrumentos presentan una adecuada bondad de ajuste, confiabilidad y correlación entre ellos. La ausencia de diferencias en ambas escalas, entre los niveles de enseñanza (primaria y secundaria) y según género permite concluir que los instrumentos logran pesquisar los constructos de forma transversal.

La Escala de Estrategias de Evaluación Formativa presenta tres subescalas que son representativas del constructo. La retroalimentación, la coevaluación, la autoevaluación y el uso de herramientas como las rúbricas se distinguen en esta escala. Esto es un aspecto importante ya que dichas estrategias serán formativas si logran orientar el progreso de los aprendizajes del estudiantado más allá de las

calificaciones (Andrade, 2023; López-Pastor y Pérez-Pueyo, 2017). En consecuencia, estos resultados son concordantes con una comprensión de la evaluación formativa que ha destacado su capacidad para describir, analizar, valorar y direccionar el progreso del aprendizaje, en tanto permite que docentes y estudiantes identifiquen brechas entre el nivel de dominio alcanzado y el nivel de dominio esperado, evidencia el desempeño del estudiantado a partir de múltiples fuentes, convierte al propio aprendizaje y al aprendizaje de los compañeros y las compañeras en objeto de reflexión personal y orienta el desempeño hacia la mejora permanente; ya sea de modo informal o espontáneo, planificado para la interacción en el aula o bien de un modo más formal o integrado al currículum (Lipnevich et al., 2016; Panadero et al., 2012; Shavelson et al., 2008).

La Escala de Calidad de la Retroalimentación para el Aprendizaje identifica tres subescalas que refieren a dimensiones de la calidad de la retroalimentación que el profesorado ofrece al estudiantado. La subescala *instrucción focalizada* señala que retroalimentar con calidad implica ofrecer comentarios que sean claros, fáciles de entender y respetuosos, considerar las características individuales del estudiantado e indicar qué acciones o elementos específicos se pueden atender para la mejora del desempeño estudiantil en la tarea. La subescala *interacción empática* enfatiza la necesidad de reconocer el impacto que podría tener la retroalimentación en la motivación del estudiantado, tanto en sus acciones como en sus afectos, y la relevancia de escuchar sus impresiones sobre el proceso de enseñanza. Finalmente, la subescala *autorregulación del desempeño estudiantil* refiere a que una retroalimentación de calidad promueve que el estudiante identifique su grado de compromiso con la mejora y reflexione sobre su propio desempeño, ofrece comentarios temporalmente próximos a la tarea para favorecer la utilidad percibida y visualiza el desempeño correcto u óptimo del estudiantado en la tarea. Estos resultados concuerdan con aspectos que la literatura destaca que son básicos para definir qué hay que retroalimentar y cómo hacerlo (Panadero y Lipnevich, 2022). Aunque la Escala de Calidad de la Retroalimentación para el Aprendizaje presenta dichas subescalas como una expresión de la calidad de la retroalimentación docente, en el ejercicio de la enseñanza el profesorado podría evaluar cuál o cuáles de dichos factores puede enfatizar, según sea el progreso del estudiante en los niveles de dominio o según sea la situación de enseñanza. De este modo, como se evidencia en los resultados de este trabajo, una retroalimentación de calidad es específica, indica qué y cómo mejorar, es respetuosa con las características individuales del estudiante, es empática con la experiencia estudiantil y consigue que el estudiantado aumente su autorregulación al invitarle a reflexionar sobre su propio desempeño en las tareas de aprendizaje (Adarkwah, 2021; Ossenberget al., 2019; Rowe, 2017; Tay y Lam, 2022).

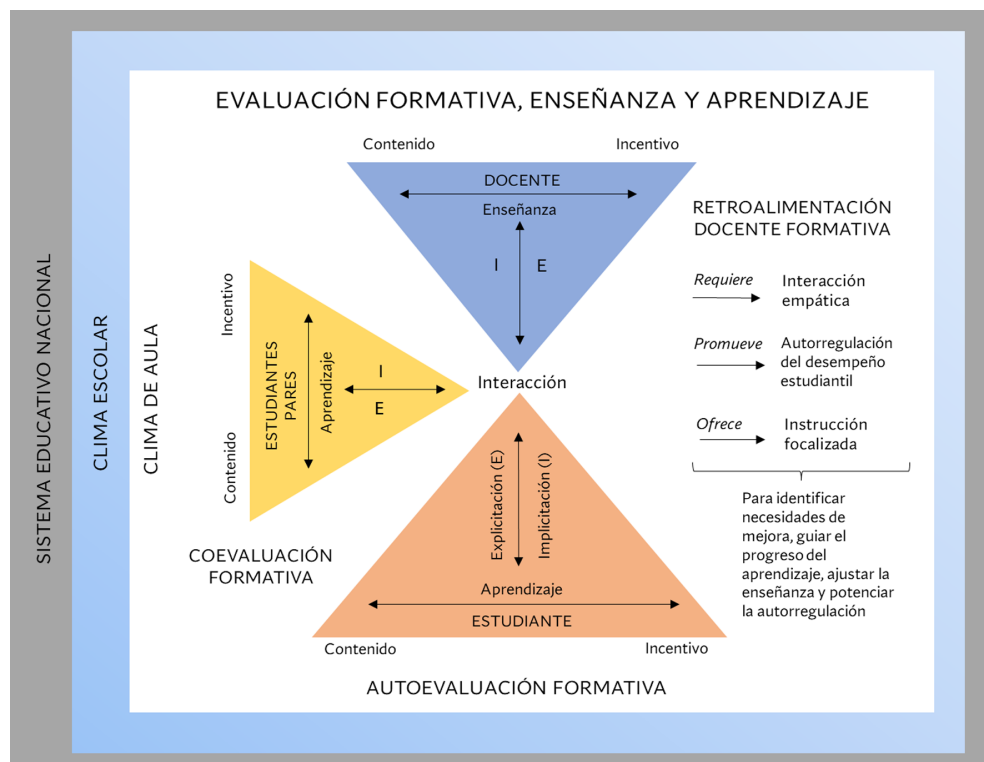
Los alcances teóricos de la presente investigación permiten visualizar la conexión entre las variables estudiadas. La Figura 3 ofrece un modelo heurístico que ilustra las relaciones entre cómo el presente trabajo ha definido a la evaluación formativa y a la calidad de la retroalimentación con los procesos del aprendizaje y la enseñanza; considerando las influencias que ejercen los climas de aula, los climas escolares y los sistemas educativos nacionales. La evaluación formativa y sus estrategias (retroalimentación, coevaluación, autoevaluación) configuran un soporte técnico y experiencial que permite hacer del propio aprendizaje y del aprendizaje de compañeros y compañeras de clase, un objeto de análisis y de práctica permanente en función de los objetivos propuestos por el profesorado. Desde aquella interacción se movilizan conceptos (contenidos) y motivaciones (incentivos) que nutren y orientan el desempeño y su sentido, tanto en el estudiantado como en el profesorado (Illeris, 2014). Los contenidos simbólicos, datos o información sobre el propio desempeño en las actividades de aprendizaje se internalizan mediante mecanismos de implicación y se exteriorizan mediante procesos de explicitación o transmisión que son inherentes a la corporalidad (Pozo, 2017). En dicho proceso, las estrategias de evaluación formativa que organiza el profesorado dirigen, incentivan y nutren -desde el contexto- el progreso en el dominio de los aprendizajes estudiantiles, ya que hacen de la retroalimentación una interacción comunicativa transformadora, ya sea a modo de diálogo intrapersonal (retroalimentación interna) o a partir del diálogo interpersonal; por sus contenidos (focalización en la tarea), por su forma (empatía) y por su mecanismo (autorregulación). Todo ello se ve favorecido por adecuados climas escolares y climas de aula (Heritage, 2010; Matthews et al., 2023; Pat-El et al., 2013). Como señalan van der Kleij y DeLuca (2023) las políticas públicas educativas pueden favorecer la realización de evaluación formativa mediante la promoción programas de formación docente que abunden en la mejora de la autonomía escolar, pero pueden obstaculizar su realización generando currículos muy extensos o no preocupándose del desgaste profesional del profesorado o por la mejora de la infraestructura escolar (van der Kleij et al., 2018).

## Limitaciones

El profesorado que ha participado de esta investigación pertenece al contexto educativo chileno, de modo que la extrapolación de estos resultados a otros contextos debe hacerse con precauciones y teniendo en consideración las similitudes socioculturales. El proceso de medición es transversal y recoge el autorreporte de docentes que mayoritariamente pertenecen a escuelas que formalmente se orientan a la calidad y a la mejora educativa. Únicamente se recogen y analizan datos de docentes lo que invita a explorar el impacto de estos constructos en medidas del estudiantado (Adarkwah, 2021).

**Figura 3**

Diagrama explicativo de la conexión entre el aprendizaje, la enseñanza y la evaluación formativa



Nota. Elaboración propia. Modelo heurístico que conecta los resultados de la presente investigación y las contribuciones de Illeris (2014) y Pozo (2017).

## Desarrollos futuros

Se propone estudiar la validez predictiva de ambas escalas con mediciones que incorporen procesos estudiantiles, de forma longitudinal, multinivel, en distintos contextos socioculturales y con una mayor cantidad de participantes. De igual modo, se sugiere continuar la investigación de estas escalas en procesos de evaluación formativa digital (Hooley y Thorpe, 2017) y en la educación superior. Asimismo, será propicio añadir otras estrategias de evaluación formativa, por ejemplo, aquellas relacionadas a la producción de evidencias de aprendizaje y la construcción compartida de los criterios de evaluación o niveles de dominio junto

al estudiantado. Aquello permitirá destacar el valor de la interacción interpersonal y pedagógica presente en la evaluación formativa, como uno de los principales motores del aprendizaje humano, dado su rol motivacional, conceptual, comportamental e identitario (Illeris, 2014; Pozo, 2017; Vygotsky, 1979).

## CONCLUSIONES

La presente investigación contribuye a avanzar en el camino hacia una metodología de evaluación formativa mejor definida, particularmente en contextos educativos escolares de primaria y secundaria, y desde el autorreporte del profesorado. Se han diseñado y validado dos escalas breves que identifican unos elementos que son centrales, tanto en la evaluación formativa como en la calidad de la retroalimentación docente. Además, ambos instrumentos permiten el autorreporte de acciones o creencias del profesorado que pueden relacionarse con la evaluación formativa y con una retroalimentación más espontánea o informal y con aquella que presenta mayores niveles de estructura (Bond et al., 2020). Ambas escalas pueden ser de utilidad para nuevas investigaciones, para la implementación de programas de asesoramiento a docentes en formación y en ejercicio (Matthews et al., 2023; Pat-El et al., 2013) y para el desarrollo de habilidades o competencias en el estudiantado (Shavelson et al., 2008).

## FINANCIACIÓN

Este trabajo ha sido financiado por la Agencia Nacional de Investigación y Desarrollo (ANID). Programa de Becas. Doctorado Becas Chile 2019 – 72200107.

## AGRADECIMIENTOS

Proyecto de Generación de Conocimiento: «Motivación. Evaluación y Autorregulación V». PID2022-138175NB-100. Ministerio de Ciencia e Innovación. España.

Centro de Investigación para la Educación Inclusiva. Programa SCIA-ANID CIE160009. Chile.

## REFERENCIAS BIBLIOGRÁFICAS

Adarkwah, M. A. (2021). The power of assessment feedback in teaching and learning: a narrative review and synthesis of the literature. *SN Social Science*, 1(75). 1-44. <https://doi.org/10.1007/s43545-021-00086-w>



- Andrade, H. L. (2023). What Is Next for Rubrics?: A Reflection on Where We Are and Where to Go From Here En C. Gonsalves, & J. Pearson (Eds.), *Improving Learning Through Assessment Rubrics: Student Awareness of What and How They Learn* (pp. 314-326). IGI Global. <https://doi.org/10.4018/978-1-6684-6086-3.ch017>
- Alqassab, M., Strijbos, J. W., & Panadero, E. (2023). A Systematic Review of Peer Assessment Design Elements. *Educational Psychology Review*, 35(18), 1-36. <https://doi.org/10.1007/s10648-023-09723-7>
- Bond, E., Woolcott, G., & Markopoulos, C. (2020). Why aren't teachers using formative assessment? What can be done about it? *Assessment Matters*, 14, 112–136. <https://doi.org/10.18296/am.0046>
- Marrón, G., & Harris, L. (2013). Student self-Assessment. En J. McMillan (Ed.), *SAGE Handbook of Research on Classroom Assessment* (pp. 367-393). SABIO. <https://doi.org/10.4135/9781452218649>
- Centro de estudios MINEDUC. (2022). *Apuntes 22: Variación en la información estadística de los docentes de la educación en desempeño, año 2021*. Centro de Estudios Ministerio de Educación de Chile (MINEDUC). <https://bit.ly/3VDJQQd>
- Cheung, G. W., & Lau, R.S. (2008). Testing Mediation and Suppression Effects of Latent Variables: Bootstrapping with Structural Equation Models. *Organizational Research Methods*, 11(2), 296–325. <https://doi.org/10.1177/1094428107300343>
- Fan, X. (2003). Using commonly available software for bootstrapping in both substantive and measurement analyses. *Educational and Psychological Measurement*, 63(1), 24-50. <https://doi.org/10.1177/0013164402239315>
- Greene, B. A., Miller, R. B., Crowson, S. M., Duque, B. L., & Akey, K. L. (2004). Predicting high school students' cognitive engagement and achievement: Contributions of classroom perceptions and motivation. *Contemporary Educational Psychology*, 29, 462–482. <https://doi.org/10.1016/j.cedpsych.2004.01.006>
- Harris, L., & Brown. G. (2022). *Self-Assessment*. Rutledge. <https://doi.org/10.4324/9781138609877-REE1-1>
- Heritage, M. (2010). *Formative assessment: Making it happen in the classroom*. <https://dx.doi.org/10.4135/9781452219493>
- Hooley, D. S., & Thorpe. J. (2017). The effects of formative reading assessments closely linked to classroom texts on high school reading comprehension. *Educational Technology Research and Development*, (65), 1215–1238. <https://doi.org/10.1007/s11423-017-9514-5>
- Hortigüela, D., Pérez-Pueyo, A., & González-Calvo, G. (2019). Pero... ¿a qué nos referimos realmente con la evaluación formativa y compartida?: Confusiones habituales y reflexiones prácticas. *Revista Iberoamericana de Evaluación Educativa*, 12(1). <https://doi.org/10.15366/riee2019.12.1.001>
- Illeris, K. (2014). Transformative Learning and Identity. *Journal of Transformative Education*, 12(2). 148–163. <https://doi.org/10.1177/1541344614548423>

- Klute, M., Apthorp, H., Harlacher, J., & Reale, M. (2017). *Formative assessment and elementary school student academic achievement: A review of the evidence* (REL 2017–259). U.S. Department of Education. Institute of Education Sciences. National Center for Education Evaluation and Regional Assistance. Regional Educational Laboratory Central. Central Laboratorio Educativo Regional.
- Lipnevich, A., & Panadero. E. (2021). A Review of Feedback Models and Theories: Descriptions, Definitions and Conclusions. *Frontiers in Education*, 6, Artículo 720195. <https://bit.ly/3vH3foO>
- Lipnevich, A., Berg, A., & Smith, J. (2016). Toward a model of student response to feedback. En G. T. Marrón., & I. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 159–185). Routledge. <https://doi.org/10.4324/9781315749136>
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica. Revisada y actualizada. *Anales de Psicología*, 30(3), 1151-1169. <https://dx.doi.org/10.6018/analesps.30.3.199361>
- López-Pastor, V., & Pérez-Pueyo. A. (2017). *Evaluación formativa y compartida en educación: experiencias de éxito en todas las etapas educativas*. Universidad de León. <https://bit.ly/4abtkeZ>
- Mardia, K. V. (1974). Applications of Some Measures of Multivariate Skewness and Kurtosis in Testing Normality and Robustness Studies. *The Indian Journal of Statistics*, 36(2), 115-128. <https://bit.ly/49eBVfE>
- Matthews, K., Sherwood, C., Enright. E., & Cook-Sather, R. (2023). What do students and teachers talk about when they talk together about feedback and assessment? Expanding notions of feedback literacy through pedagogical partnership. *Assessment y Evaluation in Higher Education*, 49(1), 26-38. [10.1080/02602938.2023.2170977](https://doi.org/10.1080/02602938.2023.2170977)
- Moos, C., & Brookhart. S. (2019). *Advancing In Every Classroom: A Guide For Instructional Leaders*. ASCD. <https://bit.ly/4cAflai>
- Ossenberg, C., Henderson, A., & Mitchell, M. (2019). What attributes guide best practice for effective feedback? A scoping reviews. *Advances in Health Sciences Education*, (24), 383–401. <https://doi.org/10.1007/s10459-018-9854-x>
- Panadero, E., & Lipnevich. R. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review*, (35). <https://doi.org/10.1016/j.edurev.2021.100416>
- Panadero, E., Alonso Tapia, J., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation. Learning and self-efficacy in secondary education. *Learning and Individual Differences*, 22(6), 806-813. <https://doi.org/10.1016/j.lindif.2012.04.007>

- Panadero, E., Alqassab, M., Fernández-Ruiz, J., & Ocampo, J. (2023) A systematic review on peer assessment: intrapersonal and interpersonal factors. *Assessment y Evaluation in Higher Education*, 8, 1053-1075. [10.1080/02602938.2023.2164884](https://doi.org/10.1080/02602938.2023.2164884)
- Pat-El, R. J., Tillema, H., Segers, M., & Vedder, P. (2013). Validation of Assessment for Learning Questionnaires for teachers and students. *The British journal of educational psychology*, 83(1), 98-113. <https://doi.org/10.1111/j.2044-8279.2011.02057.x>
- Popham., J. (2013). *Evaluación Transformativa: el poder transformador de la evaluación formativa*. Narcea.
- Pozo, J. I. (2017). Aprender más allá del cuerpo: de las representaciones encarnadas a la explicitación mediada por representaciones externas. *Infancia y Aprendizaje*, 40(2), 219-276. [10.1080/02103702.2017.1306942](https://doi.org/10.1080/02103702.2017.1306942)
- Rowe, A. D. (2017). Feelings about feedback: the role of emotions in assessment for learning. En D., Carless, S. Bridges, C. Chan, & R. Glofcheski (Eds), *Scaling up assessment for learning in higher education. The enabling power of assessment (vol 5)*. Springer.[https://doi.org/10.1007/978-981-10-3045-1\\_11](https://doi.org/10.1007/978-981-10-3045-1_11)
- Ruiz, M., Pardo, A., & San Martín. R. (2010). Modelos de ecuaciones estructurales. *Papeles del Psicólogo*, 31(1), 34-35. <https://bit.ly/3TANiIL>
- Sánchez, C. E., Atkinson, K. M., Koenka, A. C., Moshontz, H., & Cooper, H. (2017). Self-grading and peer-grading for formative and summative assessments in 3rd through 12th grade classrooms: A meta-analysis. *Journal of Educational Psychology*, 109(8), 1049-1066. <https://doi.org/10.1037/edu0000190>
- Shavelson, R. J, Young, D. B, Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A., ..., & Yin, Y. (2008). On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment developers. *Applied Measurement in Education*, 21(4), 295–314. <https://doi.org/10.1080/08957340802347647>
- Tay H. Y. & Lam, K. W. L. (2022). Students' engagement across a typology of teacher feedback practices. *Educational Research for Policy and Practice*, (21), 427–445. <https://doi.org/10.1007/s10671-022-09315-2>
- van der Kleij, F. M., & DeLuca. C. (2023). Implementation of assessment for learning. En R. Tierney, F. Rizvi, & K. Ercikan (Eds.), *International Encyclopedia of Education* (pp. 147-154). <https://doi.org/10.1016/B978-0-12-818630-5.09028-X>
- Van der Kleij, F. M., Cumming, J., & Looney, A. (2018). Policy expectations and support for teacher formative assessment in Australian education reform. *Assessment in Education: Principles. Policy y Practice*, 25(6), 620-637. [10.1080/0969594X.2017.1374924](https://doi.org/10.1080/0969594X.2017.1374924)
- Vygotsky, L. S. (1979). Interacción entre aprendizaje y desarrollo. En M., Cole, J. Vera, S. Scribner, & E. Souberman (Eds.), *El desarrollo de los procesos psicológicos superiores* (pp. 123-140). Crítica-Grijalbo.

Wulandari, D., Sutrisno, S., & Nirwana, M. B. (2021). Mardia's Skewness and Kurtosis for Assessing Normality Assumption in Multivariate Regression. *Enthusiastic: International Journal of Applied Statistics and Data Science*, 1(1), 1 – 6. <https://doi.org/10.20885/enthusiastic.vol1.iss1.art1>