


Creencias docentes sobre la evaluación y su relación con el rendimiento de los estudiantes de 4.º grado de primaria

Responsable del estudio:
Diego Hermoza Cárdenas 

En caso de consultas sobre
este artículo, escribir a:
medicion@minedu.gob.pe

Ministerio de Educación del Perú
Calle del Comercio 193, San Borja
Hecho el Depósito Legal en la
Biblioteca Nacional del Perú
N.º 2025-13171
Primera edición digital
Febrero, 2026

Resumen: Las creencias docentes sobre la evaluación son un componente central de la práctica pedagógica, pues influyen en cómo los docentes planifican el currículo, seleccionan estrategias de enseñanza y evalúan los aprendizajes de sus estudiantes. En este marco, el presente estudio analiza algunas de las creencias sobre la evaluación en el aula, recogidas en la Evaluación Nacional de Logros de Aprendizaje (ENLA) 2024 a partir del autorreporte de los docentes de estudiantes de 4.º grado de primaria. Los resultados indican que un porcentaje relevante de docentes se encuentra de acuerdo con creencias vinculadas a una visión limitada del sentido formativo de la evaluación, reflejadas en la priorización del producto final y prácticas restringidas de retroalimentación. Se identificó que dichas creencias fueron más frecuentes entre docentes con menos años de experiencia y sin título pedagógico, y que los estudiantes cuyos docentes sostienen estas creencias tienden a alcanzar un menor rendimiento académico en Lectura y Matemática. Los hallazgos resaltan la importancia de continuar con acciones de fortalecimiento de las capacidades docentes para diseñar y aplicar estrategias de evaluación desde un enfoque formativo.

Palabras clave: creencias docentes, retroalimentación, evaluación formativa, factores asociados.

Citar esta publicación de la siguiente manera:

Ministerio de Educación del Perú. (2026). *Creencias docentes sobre la evaluación y su relación con el rendimiento de los estudiantes de 4.º grado de primaria* (Estudios Breves N.º 18). Oficina de Medición de la Calidad de los Aprendizajes.

Introducción

En el ámbito educativo, la evaluación en el aula permite a los docentes conocer lo que los estudiantes saben y pueden hacer, y decidir si ese aprendizaje responde a lo esperado (Baird et al., 2017). Su valor radica en ofrecer múltiples oportunidades para recoger evidencias del aprendizaje y brindar retroalimentación oportuna que favorezca su mejora. Sin embargo, lejos de ser una actividad meramente técnica, la evaluación está cargada de concepciones sobre qué y cómo se debe aprender, así como sobre el rol del docente y del estudiante en ese proceso (Black y Wiliam, 2009; Brown, 2004). Así, qué y cómo evalúa un docente refleja sus creencias y conocimientos pedagógicos, y constituye parte de su identidad profesional (Looney et al., 2018; Ravela et al., 2017).

En el caso peruano, el Currículo Nacional de la Educación Básica (CNEB) plantea la evaluación desde un enfoque formativo, concebido como un proceso sistemático que recoge y valora información relevante acerca del nivel de desarrollo de las competencias en cada estudiante, con el fin de contribuir oportunamente a mejorar su aprendizaje (Ministerio de Educación del Perú [Minedu], 2017). En este marco, se reconoce tanto la evaluación para el aprendizaje, orientada a la retroalimentación y la mejora continua, como la evaluación del aprendizaje, de carácter certificador, pero que también cumple una función formativa (Minedu, 2020). Así, el enfoque formativo integra ambas finalidades como partes de un mismo proceso orientado al desarrollo de competencias.

Además, este énfasis no solo está presente en el currículo, sino también en las competencias y desempeños del Marco del Buen Desempeño Docente, el cual establece que la evaluación continua y la retroalimentación son componentes centrales de la labor del docente (Minedu, 2014). Esto evidencia que la retroalimentación formativa constituye un componente explícito de las expectativas pedagógicas establecidas por el sistema educativo.

Las prácticas de evaluación formativa pueden tener un impacto positivo en el aprendizaje estudiantil. Investigaciones internacionales evidencian que la retroalimentación oportuna, la evaluación continua y el involucramiento activo del estudiante favorecen la motivación, la autonomía y el rendimiento académico (Hattie y Timperley, 2007; Lee et al., 2020; Yan y Chiu, 2023). Asimismo, la naturaleza de las tareas de evaluación influye en cómo los estudiantes entienden qué es importante aprender y en su disposición a participar en el proceso educativo (Näsström et al., 2021).

A pesar de su amplio respaldo en la literatura pedagógica, es necesario señalar que la evaluación formativa representa un enfoque de alto compromiso, en la medida en que su implementación requiere que el docente desarrolle capacidades pedagógicas, diagnósticas y comunicativas. Estas incluyen definir objetivos claros, diseñar instrumentos pertinentes, interpretar evidencias y retroalimentar a los estudiantes (Brookhart, 2011; Xu y Brown, 2016). Evaluar no se reduce a aplicar técnicas específicas; requiere articular profundamente la enseñanza y la evaluación, así como utilizar evidencia para orientar decisiones pedagógicas

(Black y Wiliam, 2009; Heritage, 2007). Precisamente por su amplitud, ha sido definida de formas diversas en la práctica, lo que en ocasiones genera confusiones o simplificaciones que pueden diluir su potencial transformador (Bennett, 2011).

No obstante, en el contexto peruano aún son escasos los estudios que examinan de manera específica las creencias docentes sobre la evaluación y su vínculo con los aprendizajes de los estudiantes. Una comprensión de cómo los docentes conciben la evaluación, qué prácticas privilegian y qué sentidos les atribuyen resulta fundamental para entender las posibilidades y los límites de una implementación efectiva de la evaluación formativa. En ese sentido, utilizando los datos de la Evaluación Nacional de Logros de Aprendizaje (ENLA) 2024 de 4.º grado de primaria, el presente estudio tiene como propósito examinar las creencias de los docentes sobre la evaluación en el aula y analizar su relación con el desempeño de los estudiantes.

La siguiente sección presenta los enfoques teóricos sobre la evaluación formativa y sumativa, así como las creencias que orientan las prácticas docentes, tanto en el plano internacional como en el contexto peruano.

Evaluación formativa y sumativa

La literatura sobre el rol de la evaluación en el aula ha identificado dos finalidades principales¹ (Baird et al., 2017; Ravela, 2009; Schellekens et al., 2021): i) una función formativa, orientada a mejorar el aprendizaje; y ii) una función sumativa, centrada en certificar la competencia de los estudiantes. También se las denomina *evaluación para el aprendizaje* y *evaluación del aprendizaje*, respectivamente (Laveault y Allal, 2016; Schellekens et al., 2021). A continuación, se describen los principales elementos de ambos procesos.

La evaluación formativa busca mejorar los aprendizajes utilizando evidencias del logro de los estudiantes para decidir próximos pasos en la enseñanza (Black y Wiliam, 2009). De acuerdo con los autores, su implementación se apoya en cinco estrategias claves: 1) clarificar objetivos y criterios; 2) proponer tareas que evidencien comprensión; 3) dar retroalimentación orientada al progreso; 4) fomentar el aprendizaje entre pares; y 5) promover la autonomía del estudiante. A su vez, Brookhart (2011) subraya que una práctica formativa efectiva requiere identificar e interpretar adecuadamente el nivel de aprendizaje para ajustar la enseñanza en función de las evidencias.

La evaluación sumativa, en cambio, busca certificar lo aprendido al término de un proceso (Ravela, 2009), expresándose en calificaciones, informes o decisiones de promoción (Dolin et al., 2018). Cumple además una función de rendición de cuentas frente a estudiantes, familias, docentes y autoridades educativas. No obstante, no constituye una práctica neutral,

¹Algunos autores también discuten el rol de la evaluación diagnóstica y de la *evaluación como aprendizaje*. La primera se realiza al inicio de un ciclo o unidad para identificar conocimientos previos; la segunda enfatiza el rol del estudiante al integrar inseparablemente evaluación y aprendizaje (Laveault y Allal, 2016).

sino social, pues es influida por normas, valores y finalidades institucionales y culturales (Ravela et al., 2017). Su uso predominante, especialmente cuando se convierte en el único referente para valorar el aprendizaje y tomar decisiones pedagógicas, puede limitar el papel de la evaluación al reducirla a un mecanismo de control, restando fuerza a su potencial formativo (Ravela et al., 2017).

Asimismo, es importante señalar que no existe una completa dicotomía entre las prácticas formativas y sumativas (Black y Wiliam, 2009; Dolin et al., 2018), ya que ambas funciones pueden coexistir y articularse. Una misma actividad puede ser utilizada por el docente tanto para retroalimentar el aprendizaje como para asignar una calificación, dependiendo del propósito con que se emplee y del momento en que se integre al proceso de enseñanza.

Creencias docentes sobre la evaluación: contexto internacional

Las creencias son construcciones mentales subjetivas, pero profundamente arraigadas, las cuales pueden ser asumidas como verdades personales (Pajares, 1992). En educación, se reconoce que las creencias docentes pueden orientar su práctica, influyendo en el tipo y la calidad de las actividades que desarrollan en el aula (Buehl y Beck, 2014). En el marco de la evaluación, estas creencias aluden a los supuestos y comprensiones que los docentes consideran verdaderos respecto de qué, cómo y para qué evaluar a los estudiantes (Barnes et al., 2014).

La literatura internacional ha conceptualizado las creencias docentes sobre la evaluación como un continuo que va desde un enfoque pedagógico hasta uno centrado en la rendición de cuentas (Barnes et al., 2014; Remesal, 2011). El enfoque pedagógico concibe la evaluación como una herramienta para regular la enseñanza y promover el aprendizaje, mientras que el enfoque de rendición de cuentas es entendido como un mecanismo para cuantificar resultados y calificar tanto a los estudiantes como a la escuela. Ambas comprensiones se vinculan estrechamente con las funciones formativa y sumativa de la evaluación, respectivamente (Monteiro et al., 2021). Entre algunas aproximaciones empíricas, algunos estudios (Barnes et al., 2017; Brown, 2004) identifican estos dos enfoques sobre los propósitos de la evaluación al momento de analizar las creencias docentes en diferentes países: uno orientado a la mejora del proceso de enseñanza y aprendizaje (formativo) versus otro orientado a la rendición de cuentas (sumativo).

Asimismo, diversos estudios coinciden en que las creencias sobre la evaluación dependen fuertemente del contexto. Por ejemplo, al aplicarse el instrumento desarrollado por Brown (2004) en distintos países, se encontró que la estructura factorial del cuestionario no se replicaba de manera consistente (Brown et al., 2019). Esto sugiere que no existe un modelo global único y que las creencias docentes sobre la evaluación pueden variar según el contexto local. Asimismo, otros investigadores advierten que las creencias no siempre se reflejan en la práctica, ya que las exigencias del entorno pueden llevar a los docentes a adoptar enfoques más sumativos y centrados en la rendición de cuentas, incluso cuando valoran la evaluación formativa (Hidalgo y Murillo, 2021). En esa línea, una revisión sistemática de estudios sobre el

tema señala que diversos factores personales (como la formación, la autoeficacia o las creencias pedagógicas) y escolares (como el apoyo institucional, las políticas nacionales o las condiciones de trabajo) pueden condicionar la brecha entre la intención de aplicar prácticas de evaluación formativa y su ejecución en el aula (Yan et al., 2021).

Creencias docentes sobre la evaluación: contexto peruano

Acerca del contexto peruano, algunas investigaciones han abordado las creencias y prácticas docentes en torno a la evaluación (Eguren et al., 2019; León, 2018; Llanos y Tapia, 2021). La mayoría de estos estudios se basa en aproximaciones cualitativas a partir de observaciones de aula, análisis de cuadernos y entrevistas a docentes. Estas indagaciones han permitido identificar patrones relevantes sobre cómo se concibe y se lleva a cabo la evaluación en el aula. En particular, se han documentado creencias y prácticas relacionadas con la finalidad de la evaluación, las formas de retroalimentación y los criterios utilizados para valorar el aprendizaje.

En cuanto a la finalidad de la evaluación, los estudios identifican el predominio de una orientación hacia la certificación más que a una mejora del aprendizaje. Uno de ellos señala que los docentes tienden a confundir los fines formativos y los sumativos de la evaluación, lo cual limita sus usos para apoyar el aprendizaje; esta investigación se basa en entrevistas a docentes y análisis de cuadernos de 160 aulas de 6.º grado de primaria en ocho países latinoamericanos incluido Perú (Ravela, 2009). Un estudio posterior, realizado con docentes de secundaria de Chile, Colombia, Perú y Uruguay, se observa que las consignas de evaluación se centran principalmente en la reproducción de contenidos, con escasas oportunidades para que los estudiantes movilicen procesos cognitivos más complejos (Ravela et al., 2014). En la misma línea, otro estudio, realizado a partir de observaciones de aula y entrevistas en nueve escuelas públicas de primaria en Perú, muestra que las evaluaciones más frecuentes replican el formato de pruebas estandarizadas, reduciendo la evaluación a un ejercicio de verificación más que de una oportunidad para seguir aprendiendo (León, 2018).

En cuanto a las formas de retroalimentación, se observa que estas suelen ser poco específicas y escasamente orientadas al aprendizaje. A partir de un conjunto de observaciones de aula realizadas entre 2004 y 2013 en escuelas públicas de distintas regiones del país, una investigación concluyó que la retroalimentación docente suele limitarse a correcciones generales o expresiones de ánimo, sin ofrecer orientaciones concretas para la mejora del desempeño del estudiante (Eguren et al., 2019). Otro estudio, basado en observaciones en doce escuelas multigrado, señala que la mayoría de interacciones se reducen a determinar si las respuestas de los estudiantes son correctas o incorrectas, sin generar espacios para la elaboración de ideas ni la reflexión sobre los contenidos trabajados (Llanos y Tapia, 2021). A

ello se suman los informes del Monitoreo de Prácticas Escolares² (Minedu, 2024), los cuales reportan baja frecuencia de retroalimentación escrita o individualizada y debilidades en el monitoreo del aprendizaje.

Respecto de los criterios de evaluación, se ha identificado una tendencia a privilegiar aspectos formales o de presentación sobre procesos cognitivos involucrados. En ciertas escuelas de primaria, se valoran elementos como la caligrafía o el uso de colores, lo que refleja una concepción de la evaluación centrada en la corrección de productos acabados y no en la comprensión o el razonamiento (Eguren et al., 2019). De modo similar, se advierte que las calificaciones suelen basarse en la acumulación de actividades o comportamientos observados, sin una referencia clara a criterios pedagógicos relacionados con el aprendizaje (Ravela et al., 2014).

Los estudios cualitativos permiten visibilizar desafíos importantes para la implementación efectiva de la evaluación formativa en las escuelas peruanas. En general, se observa que persisten creencias y prácticas que limitan su uso como herramienta pedagógica. Por ejemplo, la evaluación formativa tiende a ser concebida principalmente como una estrategia para promover aspectos actitudinales o de formación integral (por ejemplo, la responsabilidad o la puntualidad en la entrega de trabajos), más que como un medio para desarrollar competencias cognitivas (Ravela et al., 2014). También se señala que la evaluación formativa está prácticamente ausente en las aulas observadas (León, 2018).

Preguntas de investigación

Este panorama pone en evidencia la necesidad de estudios cuantitativos que complementen las aproximaciones cualitativas previas, proporcionando una visión más sistemática y generalizable sobre las creencias docentes y su relación con el desempeño de los estudiantes. Para ello, se plantearon las siguientes preguntas de investigación:

- ¿Cuáles creencias sobre la evaluación, que se distancian de un enfoque formativo, son manifestadas por los docentes de 4.º grado de primaria?
- ¿Cómo varían las creencias docentes sobre la evaluación según características individuales (como sexo, título pedagógico y experiencia profesional) y escolares (gestión y área de la escuela)?
- ¿Cómo se relacionan las creencias docentes sobre la evaluación con el rendimiento de los estudiantes de 4.º grado de primaria en Lectura y Matemática?

²El Monitoreo de Prácticas Escolares es un proceso de recopilación de información que realiza el Minedu mediante la observación de sesiones de clase en escuelas de Educación Básica Regular (inicial, primaria y secundaria). En cada aula observada, se revisan además dos cuadernos de estudiantes seleccionados al azar. Según sus reportes, solo una proporción mínima de docentes logra realizar un monitoreo continuo del progreso de los estudiantes y ofrecer retroalimentación (oral o escrita) que oriente cómo mejorar su aprendizaje.

Método

Participantes

Los datos provienen de la Evaluación Nacional de Logros de Aprendizaje (ENLA) 2024 del Minedu, la cual tuvo carácter censal ($N = 530\,367$) en 4.º grado de primaria. Además de las pruebas, la ENLA 2024 incluyó la aplicación de un cuestionario dirigido a docentes que enseñaban Comunicación y Matemática a los estudiantes evaluados. En total, participaron 27 236 docentes de 4.º grado de primaria, pertenecientes a 18 111 instituciones educativas. Estos docentes representan aproximadamente al 73 % del total de docentes de 4.º grado a nivel nacional y enseñan al 96 % de los estudiantes de dicho grado, dado que el marco poblacional de la ENLA excluye a las instituciones educativas con menos de cinco estudiantes matriculados en 4.º grado. En la tabla 1 se detalla la distribución de los docentes según sus características y el estrato de la escuela.

Tabla 1

Datos sociodemográficos de los docentes

Variable		N	%
<i>Características del docente</i>			
Sexo	Mujer	20174	74,1
	Hombre	7061	25,9
Años de experiencia	5 años o menos	3665	13,7
	Entre 6 y 20 años	12119	45,3
	Más de 20 años	10971	41,0
Título pedagógico	Sí	23129	88,9
	No	2881	11,1
<i>Estrato de la escuela</i>			
Gestión	Pública	19506	71,6
	Privada	7730	28,4
Área geográfica	Urbana	19442	71,4
	Rural	7794	28,6

Medidas

A continuación, se presentan las variables utilizadas en el estudio, junto con sus definiciones operacionales.

Creencias docentes distantes del enfoque formativo. Estas se midieron mediante un conjunto de enunciados orientados a recoger algunas de las creencias docentes sobre distintos aspectos de la evaluación en el aula. Los ítems fueron elaborados a partir de referentes conceptuales y empíricos desarrollados en el contexto peruano sobre la práctica pedagógica y la función formativa de la evaluación (Eguren et al., 2019; León, 2018; Ravela et al., 2014, 2017), y complementados con hallazgos de las salidas de campo realizadas por el equipo de análisis de la Oficina de Medición de la Calidad de los Aprendizajes (UMC) del Minedu. La

escala se organizó en tres dimensiones³. La primera dimensión, denominada *criterios de evaluación*, alude a una forma de entender la evaluación centrada principalmente en el producto final (por ejemplo, el orden o la presentación de la tarea) y a un uso de los criterios desvinculado del aprendizaje, concebidos más como requisitos formales que como herramientas para orientar la mejora del estudiante. La segunda, *evidencia del progreso*, hace referencia a una comprensión del proceso evaluativo enfocada en la calificación o la medición formal, bajo una lógica comparativa o acumulativa que soslaya la función formativa. Por último, la dimensión de *retroalimentación* recoge creencias que la reducen a momentos puntuales (como después de una prueba escrita o al final de la clase) y minimizan su valor como recurso para la mejora del aprendizaje.

Los enunciados fueron evaluados mediante escalas tipo Likert de cuatro opciones (Totalmente en desacuerdo, En desacuerdo, De acuerdo, Totalmente de acuerdo). Las tres dimensiones del cuestionario (criterios de evaluación, evidencia del progreso y retroalimentación) están conceptualmente relacionadas y reflejan un constructo general sobre la orientación de los docentes hacia la evaluación. Por ello, se estimó un modelo factorial confirmatorio de segundo orden, que sintetiza las tres dimensiones en un puntaje global sin perder la información específica de cada una. El modelo mostró un ajuste adecuado (CFI = 0,978; TLI = 0,972; RMSEA = 0,075 IC 90 % [0,074–0,077]; SRMR = 0,051). La confiabilidad del puntaje global, estimada mediante el coeficiente omega jerárquico⁴, fue de 0,80; mientras que la de cada dimensión fue de 0,79 para criterios de evaluación, 0,63 para evidencia del progreso y 0,71 para retroalimentación. El análisis se realizó con el paquete *lavaan* (Rosseel, 2012) en R (R Core Team, 2025).

Características sociodemográficas y educativas del docente. Se recogió información sobre el sexo (0 = hombre, 1 = mujer), experiencia profesional (0 = 5 años o menos, 1 = entre 6 y 20 años, 2 = más de 20 años) y tenencia de título pedagógico (0 = no, 1 = sí).

Índice socioeconómico del estudiante (ISE). Índice elaborado a partir de cinco indicadores: nivel educativo de los padres, materiales de construcción de la vivienda, servicios básicos en el hogar, tenencia de activos y otros servicios en el hogar (Minedu, 2018). Esta información fue recogida mediante un cuestionario de factores asociados dirigidos a las familias de los estudiantes. Un mayor puntaje en el ISE indica que los estudiantes provienen de familias con mayor nivel socioeconómico, relativo a sus pares con menos puntajes. Asimismo, se calculó el índice socioeconómico promedio de la escuela (ISEP), el cual representa la media simple entre todos los estudiantes evaluados pertenecientes a la misma institución educativa.

Rendimiento en Lectura y Matemática. Se consideraron los resultados de los estudiantes en las pruebas de Lectura y Matemática de la ENLA 2024. Dichos puntajes fueron estimados

³El reporte técnico de ENLA 2024 (Minedu, 2025) presenta esta escala con dos dimensiones. Este estudio, sin embargo, analiza un modelo alternativo de tres dimensiones y una estructura factorial de segundo orden.

⁴Este coeficiente estima la proporción de varianza atribuible exclusivamente al factor de segundo orden, separando la varianza de las dimensiones específicas y el error de medición (Flora, 2020).

mediante un modelo Rasch y se transformaron a una escala con media 500 y desviación estándar 100 (ver Reporte técnico de la ENLA 2024; Minedu, 2025).

Estrato de la escuela. Se recodificaron las variables de gestión y ubicación geográfica de la escuela en una sola variable categórica con tres niveles (0 = privada, 1 = pública rural y 2 = pública urbana). Esta estrategia busca simplificar los modelos y facilitar una interpretación más directa de los resultados⁵.

Análisis de datos

En primer lugar, se realizó un análisis descriptivo de las respuestas a los ítems de la escala de creencias docentes distantes del enfoque formativo, con el fin de obtener una visión general del estado de dichas creencias a nivel nacional. Este análisis se realizó por separado para cada una de las tres dimensiones —criterios de evaluación, evidencia del progreso y retroalimentación—, dado que cada una permite identificar patrones diferenciados en las creencias de los docentes.

En segundo lugar, se utilizó el puntaje global de la escala de creencias, obtenido a partir del modelo factorial de segundo orden que sintetiza las tres dimensiones del cuestionario. Con este puntaje se calculó el promedio según las características demográficas y educativas del docente, así como el estrato de la escuela a la que pertenecen. La decisión de usar el puntaje global de la variable responde tanto a la alta correlación entre los factores que la conforman, como al objetivo de facilitar la interpretación de los resultados. Además de las comparaciones de promedios entre grupos, se estimó una regresión múltiple para examinar el efecto conjunto de las características docentes y del contexto escolar sobre las creencias.

Finalmente, se estimaron una serie de modelos multinivel de dos niveles (estudiantes agrupados en escuelas) para examinar la relación entre las creencias docentes y el rendimiento de los estudiantes en Lectura y Matemática. Para este análisis, se calculó un puntaje promedio de creencias por escuela. Se utilizó una secuencia de modelos: un modelo nulo para conocer la distribución de la varianza del rendimiento, un modelo con la variable de creencias, otro modelo con controles asociados a características del docente y un modelo final que incluyó variables del contexto escolar (ISEP, tipo de gestión y área). Esta secuencia permite observar cómo cambia la magnitud del efecto de las creencias docentes al incorporar sucesivamente variables de control. Además, se estimaron modelos separados por estrato escolar (privadas, públicas rurales y públicas urbanas) para explorar la consistencia de los resultados. Los modelos se ajustaron con el paquete *lme4* (Bates et al., 2025) en R (R Core Team, 2025). Dado que los datos son censales y que con grandes muestras la significancia estadística⁶ pierde

⁵No se consideró el estrato privado rural, debido a que la proporción de estudiantes en instituciones de este estrato es muy pequeña. En la ENLA 2024, este grupo representó el 0.9 % del total de estudiantes.

⁶Aunque los datos son censales, en tablas y figuras se reportan errores estándar, los cuales solo expresan la incertidumbre de las estimaciones y no corresponden a error muestral.

relevancia práctica, el análisis se centró en la magnitud de las relaciones, complementada con tamaños del efecto⁷ (Marsh et al., 2009).

Resultados

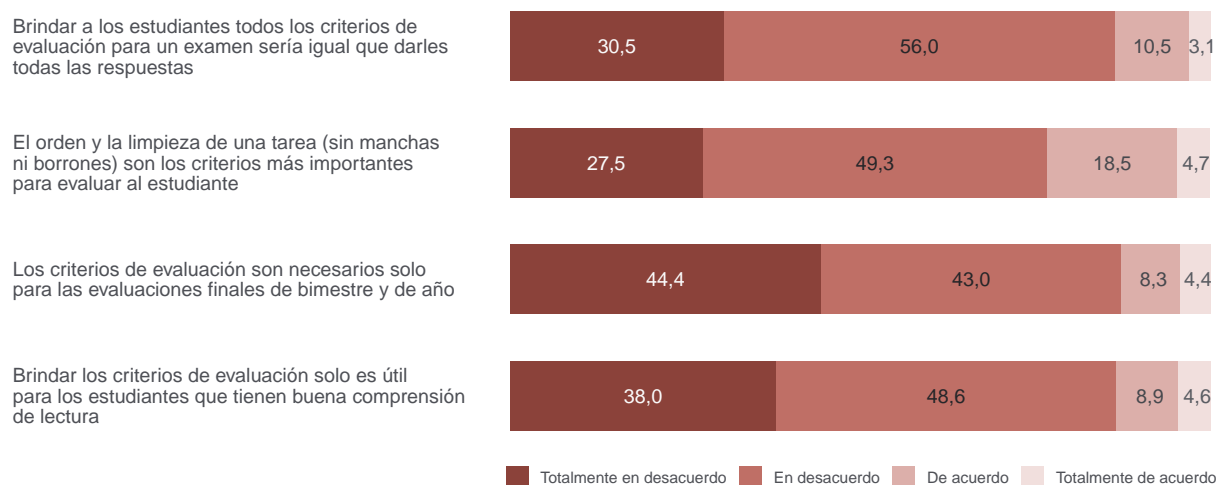
A continuación, se presentan los resultados del estudio. En primer lugar, se muestran los resultados descriptivos de la escala de creencias distantes del enfoque formativo. En segundo lugar, se presentan las diferencias en las creencias según las características de los docentes y el estrato de la escuela. Finalmente, se detallan los resultados de los modelos multinivel que examinan la relación entre las creencias docentes y el rendimiento de los estudiantes.

Resultados descriptivos

En la figura 1, se presentan las respuestas de los docentes de 4.º grado de primaria a los enunciados que conforman la dimensión de “criterios de evaluación”. En términos generales, se observa un amplio desacuerdo con las afirmaciones que forman parte de la escala. Más del 85 % de los docentes manifestó su desacuerdo con la idea de que proporcionar los criterios de evaluación equivale a “dar todas las respuestas”, que su utilidad se limita a estudiantes con buena comprensión lectora, o que estos solo son necesarios al final del bimestre. En cambio, para el enunciado referido al orden y la limpieza como criterios de evaluación, el desacuerdo fue menor (76,3 %), lo que implica que aproximadamente uno de cada cuatro docentes estuvo de acuerdo con que estos aspectos son los más importantes para evaluar al estudiante.

Figura 1

Porcentaje de respuesta de docentes de 4.º grado de primaria a la dimensión de criterios de evaluación

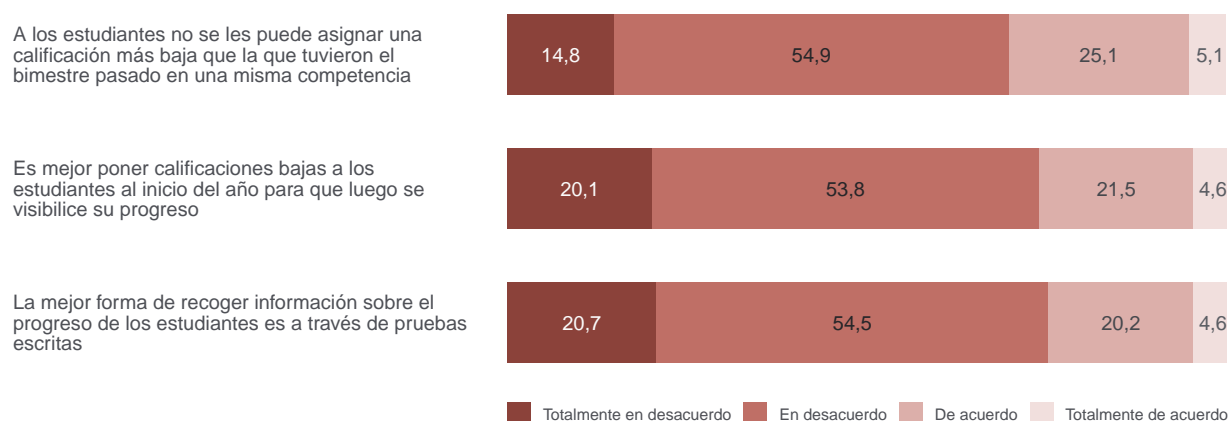


⁷El tamaño del efecto (ES) se calculó con la siguiente fórmula $ES = \frac{2 \times b \times SD_x}{\sigma_e}$ donde b es el coeficiente no estandarizado, SD_x es la desviación estándar del predictor y σ_e es desviación estándar de los residuos (el error no explicado) en el nivel de los estudiantes. Para mayor detalle, revisar Marsh et al. (2009).

En relación con los porcentajes de respuesta de la dimensión de “evidencia del progreso”, estos se presentan en la figura 2. A diferencia de lo observado en la figura 1, en esta dimensión se observa un mayor acuerdo con los enunciados. Así, aproximadamente el 30,2 % de los docentes manifestó estar de acuerdo con que no se puede asignar una calificación más baja que la obtenida previamente, mientras que un 26,1 % consideró conveniente asignar calificaciones bajas al inicio del año. Asimismo, un 24,8 % coincidió en que las pruebas escritas constituyen la mejor forma de recoger información sobre el progreso de los estudiantes.

Figura 2

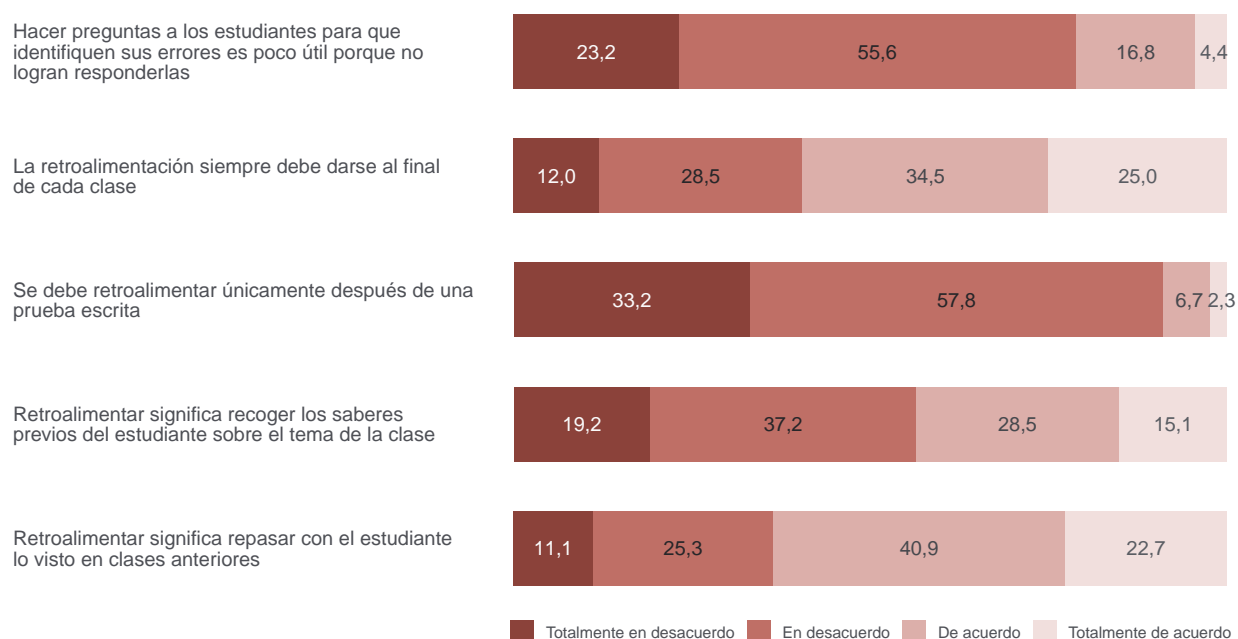
Porcentaje de respuesta de docentes de 4.º grado de primaria a la dimensión de evidencia del progreso



En la figura 3, se presentan los porcentajes de respuesta de los docentes en la dimensión de “retroalimentación”. Cerca del 60 % de los docentes manifestó estar de acuerdo o totalmente de acuerdo con una idea restringida de la retroalimentación, ya sea al señalar que esta debe darse siempre al final de cada clase (59,5 %) o que se limita a repasar lo visto en sesiones anteriores (63,6 %). Por otro lado, solo un 9,1 % de los docentes consideró que la retroalimentación debe realizarse únicamente después de una prueba escrita.

Figura 3

Porcentaje de respuesta de docentes de 4.º grado de primaria a la dimensión de retroalimentación



Relación entre las creencias distantes del enfoque formativo y las características del docente y la escuela

En la tabla 2, se presentan los puntajes estandarizados promedio ($M = 0$) y la desviación estándar ($DE = 1$) de la variable de creencias docentes distantes del enfoque formativo. Es necesario enfatizar que puntajes más altos reflejan una visión más restringida y centrada en aspectos parciales de la evaluación. Los resultados se agrupan según características individuales del docente y estrato de la escuela. A nivel individual, se observa que los docentes hombres registran un promedio superior al de las mujeres en puntajes sobre sus creencias. Asimismo, los docentes con cinco años o menos de experiencia presentan niveles más altos en este tipo de creencias en comparación con aquellos con mayor tiempo en la docencia. En cuanto a la formación, los docentes sin título pedagógico obtienen un puntaje promedio más elevado en comparación con quienes sí lo tienen. Respecto del estrato de la escuela, los docentes de escuelas privadas evidencian un promedio más alto, seguido por los de escuelas públicas rurales y escuelas públicas urbanas. En general, se observan diferencias importantes entre los grupos analizados; por ejemplo: la diferencia en la media entre hombres y mujeres es de 0,31, lo que representa aproximadamente el 30 % de una DE .

Tabla 2

Puntaje promedio de la variable de creencias distantes del enfoque formativo, según características del docente y de la escuela

		Creencias distantes del enfoque formativo
		<i>M (DE)</i>
<i>Características del docente</i>		
Sexo	Mujer	-0,08 (0,97)
	Hombre	0,22 (1,05)
Años de experiencia	5 años o menos	0,24 (0,93)
	6 a 20 años	-0,06 (1,00)
	Más de 20 años	-0,03 (1,01)
Título pedagógico	Sí	-0,04 (1,01)
	No	0,33 (0,87)
<i>Características de la escuela</i>		
Estrato	Privada	0,22 (0,87)
	Pública rural	0,05 (1,05)
	Pública urbana	-0,18 (1,02)

Además de comparar los promedios de las creencias entre grupos definidos por características docentes y del contexto escolar, se estimaron modelos de regresión para analizar simultáneamente cómo estas características se relacionan con las creencias. Los resultados se presentan en la tabla 3. En el modelo (1), que considera solo las variables individuales, se replican los patrones observados en la tabla 2: los docentes con mayor acuerdo con los enunciados que reflejan una visión limitada de la evaluación son, principalmente, hombres, sin título pedagógico o tienen menor experiencia profesional. En el modelo (2), que incorpora, además, el estrato de la escuela (con las escuelas privadas como categoría de referencia), las asociaciones de las características individuales se mantienen, aunque algunos coeficientes se reducen, como en el caso de la tenencia del título pedagógico y la experiencia mayor a 20 años. Esto sugiere que parte de la relación entre las características individuales y las creencias docentes podría estar moderada por la gestión y el área de la escuela, lo que indica que el contexto escolar contribuye a explicar la magnitud de estas asociaciones.

Tabla 3

Relación entre las creencias distantes del enfoque formativo y características individuales y escolares

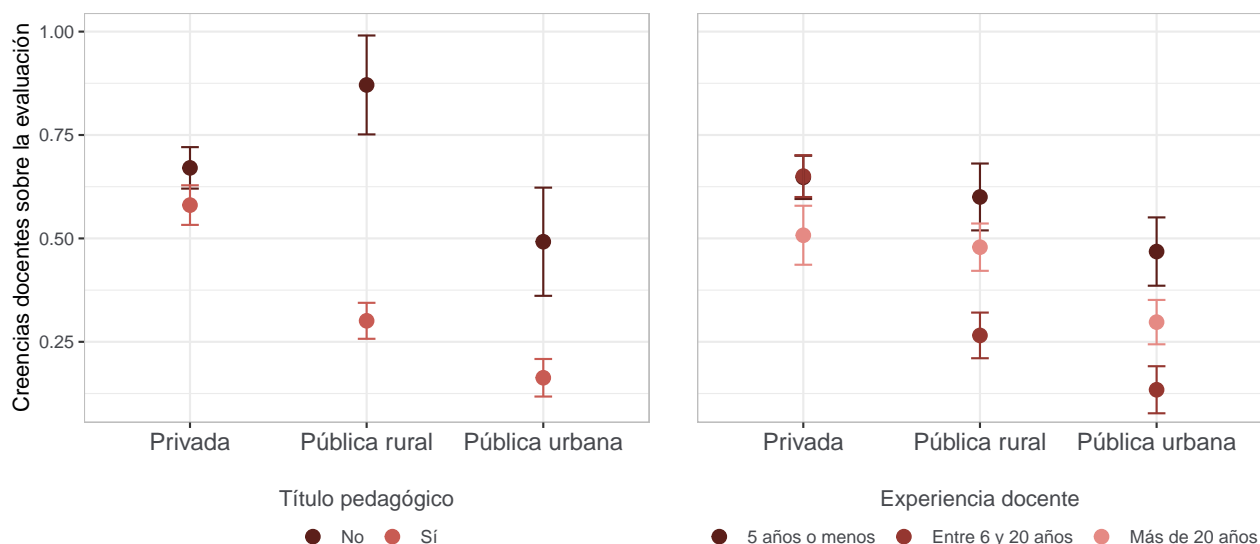
	Modelo 1	Modelo 2	Modelo 3	Modelo 4
Intercepto	0,69 (0,02)	0,73 (0,02)	0,67 (0,03)	0,67 (0,03)
Sexo (Mujer)	-0,31 (0,01)	-0,33 (0,01)	-0,33 (0,01)	-0,34 (0,01)
Título pedagógico (Sí)	-0,34 (0,02)	-0,18 (0,02)	-0,09 (0,03)	-0,19 (0,02)
Años de experiencia (6 a 20 años)	-0,21 (0,02)	-0,18 (0,02)	-0,17 (0,02)	0,00 (0,03)
Años de experiencia (más de 20 años)	-0,17 (0,02)	-0,06 (0,02)	-0,06 (0,02)	-0,13 (0,04)
Escuela pública urbana (PU)		-0,24 (0,02)	0,20 (0,06)	-0,04 (0,04)
Escuela pública rural (PR)		-0,39 (0,02)	-0,18 (0,07)	-0,18 (0,04)
Título pedagógico (Sí) × Escuela PR			-0,48 (0,07)	
Título pedagógico (Sí) × Escuela PU			-0,24 (0,07)	
Años de experiencia (6 a 20 años) × Escuela PR				-0,34 (0,05)
Años de experiencia (más de 20 años) × Escuela PR				0,02 (0,05)
Años de experiencia (6 a 20 años) × Escuela PU				-0,34 (0,05)
Años de experiencia (más de 20 años) × Escuela PU				-0,03 (0,05)
R ²	0,04	0,06	0,06	0,06
Adj. R ²	0,04	0,06	0,06	0,06
Num. obs.	25497	25497	25497	25497

De manera adicional, se estimaron dos modelos que incluyen interacciones entre el estrato de la escuela con el título pedagógico y los años de experiencia, respectivamente. Los resultados de estos modelos se presentan de manera gráfica para facilitar su interpretación (ver figura 4). Allí, se observa que los docentes con título pedagógico presentan valores promedio más bajos en la variable de creencias en comparación con aquellos que no cuentan con dicho título. Esta diferencia se amplía en las escuelas públicas, donde los docentes con título pedagógico presentan valores aún más bajos, especialmente en el área urbana. En relación con los años de experiencia docente, en las escuelas privadas, los puntajes promedio son relativamente similares entre los distintos grupos de experiencia. En cambio, en las escuelas

públicas, los docentes que tienen entre 6 y 20 años de experiencia presentaron un menor puntaje promedio; este patrón es más notorio en las escuelas urbanas. En otras palabras, los docentes de escuelas públicas urbanas con 6 a 20 años de experiencia tienden a mostrar respuestas más consistentes con una visión formativa de la evaluación.

Figura 4

Puntajes promedio estimados de creencias distantes del enfoque formativo según título pedagógico, experiencia docente y estrato de la escuela



Nota. Los puntajes promedio corresponden a estimaciones de modelos de regresión lineal. Se usaron modelos separados para las interacciones entre estrato y título pedagógico, así como entre estrato y experiencia docente, correspondientes a los modelos (3) y (4) de la tabla 3, respectivamente.

Relación entre las creencias docentes distantes del enfoque formativo y el rendimiento de los estudiantes

Los resultados de los modelos multinivel se presentan en las tablas 4 y 5, correspondientes a las áreas de Lectura y Matemática, respectivamente. Cada tabla incluye cuatro modelos: el modelo (1), sin variables predictoras (modelo nulo); el modelo (2), que incorpora las creencias docentes sobre la evaluación; el modelo (3), que añade controles a nivel docente (sexo, experiencia docente y título pedagógico); y el modelo (4), que incluye controles del contexto escolar (gestión y área e ISEP). En ambas tablas, se muestran únicamente los coeficientes asociados a las creencias docentes, dado que constituyen la variable de interés en este análisis. Los resultados completos de las regresiones, que incluyen las demás variables y parámetros de ajuste, se presentan en el Anexo.

El modelo (1) permitió estimar la proporción de varianza atribuible a las diferencias entre escuelas. En ambas áreas, algo más de una cuarta parte de la varianza total se localizó a nivel escolar (27,9% en Lectura y 26,8% en Matemática). En el modelo (2), al incorporar las creencias distantes del enfoque formativo, se observó una relación negativa en ambas áreas (Lectura: $\beta = -7,80$; Matemática: $\beta = -11,56$); es decir, los estudiantes cuyos docentes presentan

creencias más restrictivas o parciales sobre la evaluación obtienen, en promedio, entre 8 y 12 puntos menos en las pruebas de Lectura y Matemática. Estos modelos presentaron una mejora significativa en el ajuste respecto de los modelos nulos, aunque la reducción en la varianza entre escuelas fue pequeña, aproximadamente 3 % en Lectura y 4 % en Matemática.

En el modelo (3), al controlar por características individuales de los docentes (sexo, experiencia y título pedagógico), la relación negativa entre las creencias docentes sobre la evaluación y el rendimiento se mantuvo en ambas áreas, aunque con una ligera reducción en la magnitud del coeficiente (Lectura: $\beta = -7,43$; Matemática: $\beta = -10,91$). Al incorporar, en el modelo (4), las características contextuales de las escuelas (gestión y área e ISEP), la asociación negativa persistió, aunque nuevamente con una disminución en la magnitud (Lectura: $\beta = -5,64$; Matemática: $\beta = -8,24$). Estos resultados sugieren que, aun considerando diferencias entre las características docentes y escolares, los docentes que conciben la evaluación principalmente como una herramienta para calificar, con criterios que no reflejan los niveles de logro de las competencias y una retroalimentación poco formativa, tienen estudiantes con menores niveles de rendimiento.

Tabla 4

Relación entre las creencias distantes del enfoque formativo y rendimiento en Lectura

	Modelo (1) (Nulo)	Modelo (2) (Solo creencias)	Modelo (3) (Controles individuales)	Modelo (4) (Controles individuales y contextuales)
Creencias docentes		-7,52 (0,38)	-7,13 (0,38)	-5,54 (0,27)
Componentes de varianza				
Varianza residual	4676	4676	4672	4675
Varianza entre escuelas	1813	1764	1640	602
Proporción de varianza explicada				
Nivel de estudiante		0,00	0,00	0,00
Nivel de escuela		0,03	0,10	0,67

Nota. El modelo 1 no cuenta con variables predictoras y sirve de referencia para los modelos sucesivos. El modelo 2 incorpora únicamente las creencias docentes sobre la evaluación. El modelo 3 añade controles individuales del docente (sexo, título pedagógico y años de experiencia). El modelo 4 incluye, además de los controles individuales, la gestión y área de la escuela, y el ISEP. Se muestra el coeficiente estimado de la variable creencias docentes sobre la evaluación en cada modelo. Las tablas completas se presentan en la tabla A.1 en el anexo.

Tabla 5*Relación entre las creencias distantes del enfoque formativo y rendimiento en Matemática*

	Modelo (1) (Nulo)	Modelo (2) (Solo creencias)	Modelo (3) (Controles individuales)	Modelo (4) (Controles individuales y contextuales)
Creencias docentes		-11,19 (0,49)	-10,53 (0,50)	-8,09 (0,40)
Componentes de varianza				
Varianza residual	8343	8343	8339	8341
Varianza entre escuelas	3060	2952	2813	1511
Proporción de varianza explicada				
Nivel de estudiante		0,00	0,00	0,00
Nivel de escuela		0,04	0,08	0,50

Nota. El modelo 1 no cuenta con variables predictoras y sirve de referencia para los modelos sucesivos. El modelo 2 incorpora únicamente las creencias docentes sobre la evaluación. El modelo 3 añade controles individuales del docente (sexo, título pedagógico y años de experiencia). El modelo 4 incluye, además de los controles individuales, la gestión y área de la escuela, y el ISEP. Se muestra el coeficiente estimado de la variable creencias docentes sobre la evaluación en cada modelo. Las tablas completas se presentan en la tabla A.1 en el anexo.

Finalmente, se exploró si la relación entre las creencias docentes que distan de un enfoque formativo y el rendimiento de los estudiantes se mantenía al analizar por separado los distintos tipos de escuelas (privadas, públicas rurales y públicas urbanas). Para ello, se estimaron regresiones multinivel independientes para cada subgrupo, controlando por el sexo, la experiencia y el título pedagógico del docente, así como por el ISEP. La tabla 6 presenta los coeficientes estimados a partir de estos modelos⁸. En Lectura, los coeficientes se encontraron entre -5,36 y -5,65, mientras que en Matemática los coeficientes fueron ligeramente mayores, con valores entre -6,87 y -8,82. Estos resultados refuerzan la evidencia de que, independientemente del tipo de escuela, las creencias docentes con una visión limitada de la evaluación formativa se asocian con un menor logro de aprendizajes.

Tabla 6*Relación entre las creencias distantes del enfoque formativo y el rendimiento según estrato de la escuela*

Área	Privada	Pública rural	Pública urbana
Lectura	-5,36 (0,43)	-5,56 (0,46)	-5,65 (0,51)
Matemática	-6,87 (0,61)	-8,82 (0,69)	-8,05 (0,81)

Nota. Los coeficientes se estimaron a partir de un modelo multinivel que controla por sexo del docente, experiencia docente, tenencia de título pedagógico y el ISEP. Se muestra el coeficiente estimado de la variable creencias docentes sobre la evaluación en cada modelo. Las tablas completas se presentan en la tabla A.2 en el anexo.

⁸En todos los casos, los modelos que incluyen las creencias docentes mostraron un mejor ajuste respecto de los modelos nulos, lo que sugiere que estas creencias contribuyen a explicar las diferencias en el rendimiento al interior de cada estrato.

Para valorar la magnitud de las relaciones, se calcularon tamaños del efecto siguiendo el procedimiento propuesto por Marsh et al. (2009). En los modelos generales, los tamaños del efecto para las creencias docentes se encontraron entre -0,15 (cuando se incluyeron todos los controles) y -0,22 (sin controles). Al analizar por estrato, los tamaños del efecto se mantuvieron entre -0,13 y -0,17 en Lectura, y -0,13 y -0,20 en Matemática. Estos valores se consideran pequeños a moderados en el contexto de intervenciones educativas, aunque son relevantes, ya que reflejan una relación sistemática de las creencias docentes sobre los logros de los estudiantes a nivel escolar (Kraft, 2020). Por ejemplo, a nivel nacional, los estudiantes que asisten a escuelas con docentes y que presentan mayores niveles en estas creencias mostraron, en promedio, una reducción de 0,16 desviaciones estándar en sus puntajes de Matemática, incluso al considerar el contexto socioeconómico de la escuela.

Discusión

El presente estudio tuvo como objetivo examinar las creencias de los docentes acerca de la evaluación. En primer lugar, se presentaron los patrones de respuesta de los docentes, que permitieron identificar las ideas que ellos tienen sobre distintos aspectos de la evaluación. Luego, se presentaron las variables individuales del docente y del estrato escolar que se asocian a estas. Por último, se analizó la relación entre las creencias y el rendimiento de los estudiantes en Lectura y Matemática.

Los resultados descriptivos muestran que la mayoría de docentes de 4.º grado de primaria está en desacuerdo con afirmaciones centradas en una visión restringida de la evaluación. Por ejemplo, un 87 % está en desacuerdo con limitar los criterios de evaluación únicamente a las pruebas finales, y un porcentaje similar rechaza la idea de que estos criterios solo sean útiles para los estudiantes con buena comprensión lectora. Esto sugiere que, al menos declarativamente, los docentes tienden a distanciarse de enfoques más restrictivos sobre la evaluación, lo que constituye una condición clave para transitar hacia prácticas más alineadas con un enfoque formativo, que enfatiza el uso continuo de la información para apoyar el aprendizaje y la toma de decisiones pedagógicas basadas en evidencia (Black y Wiliam, 2009).

No obstante, las respuestas también evidencian una noción restringida sobre cómo recoger información acerca del aprendizaje, pues un 25 % de los docentes considera que las pruebas escritas son la mejor forma de hacerlo. Si bien este tipo de instrumentos cumplen un papel relevante en determinados contextos y aprendizajes, se debe recurrir a diversas estrategias e instrumentos, seleccionados según la naturaleza de los aprendizajes que se busca evidenciar y el momento del proceso. En un enfoque formativo, la variedad y pertinencia de las evidencias resultan fundamentales para acompañar el aprendizaje y ofrecer retroalimentación oportuna que oriente la mejora.

Del mismo modo, la creencia de que no se debe otorgar una calificación menor que la obtenida previamente, respondida por más del 30 % de los docentes, desvirtúa la función de la

evaluación como medio para identificar avances y dificultades reales en el aprendizaje. Esta práctica podría estar asociada a presiones institucionales o exigencias externas (Hidalgo y Murillo, 2021), pero también a comprensiones erróneas sobre lo que representa el progreso de los estudiantes (Ravela et al., 2014). Esto último es uno de los aspectos que discuten Dolin et al. (2018), quienes señalan que un enfoque formativo efectivo requiere que los docentes comprendan con claridad las progresiones de aprendizaje y traduzcan esas metas en criterios precisos y compartidos.

Por otra parte, se encontró que casi una cuarta parte de los docentes considera el orden y la limpieza como los criterios más importantes para evaluar. Ello sugiere que, para un sector importante, la evaluación sigue entendida como un mecanismo de control de aspectos formales más que como una oportunidad para valorar aprendizajes significativos. Estos hallazgos coinciden con los resultados de Eguren et al. (2019), quienes encontraron que, en la corrección de trabajos o tareas, los docentes suelen centrarse en elementos de forma, como los colores, el tamaño de la letra o la presentación general. Este tipo de criterios se enfocan en la apariencia del producto y no permiten evidenciar los procesos cognitivos ni las estrategias que posibilitan el aprendizaje. Desde un enfoque formativo, instrumentos de evaluación como las rúbricas buscan justamente superar esta limitación al integrar distintos criterios articulados entre sí, los cuales permiten observar los niveles de logro de una competencia más allá del resultado final o de la presentación del trabajo.

También se evidenciaron comprensiones parciales sobre la retroalimentación. Un alto porcentaje de docentes indicó que esta debe realizarse únicamente al final de la clase o limitarse a repasar contenidos vistos previamente, lo cual revela una comprensión limitada de la función reguladora de la retroalimentación en el aprendizaje. Estas prácticas se alejan de un enfoque formativo porque no necesariamente parten de la evidencia del desempeño del estudiante ni ofrecen información que le permita mejorar durante el proceso. En este sentido, como advierte Bennett (2011), la amplitud con que suele definirse la evaluación formativa favorece interpretaciones diversas y, en ocasiones, superficiales de sus componentes. Así, prácticas de retroalimentación que deberían orientar la reflexión del estudiante y apoyar su autorregulación terminan convirtiéndose en un concepto ambiguo, que cada docente interpreta a su modo y que en la práctica pierde su valor formativo (Hattie y Timperley, 2007).

En relación con las características del docente, se observaron mayores niveles de adhesión a creencias restringidas sobre la evaluación entre quienes no cuentan con título pedagógico. Estos resultados son consistentes con la idea de que la formación pedagógica inicial cumple un rol relevante en la construcción de formas de entender la evaluación. En el contexto peruano, los docentes sin título pedagógico suelen presentar dificultades para brindar retroalimentación mediante repreguntas y breves explicaciones (González et al., 2022). Así, no contar con un título pedagógico implicaría tener trayectorias formativas sin espacios sistemáticos de reflexión sobre el sentido pedagógico de la evaluación, así como sobre los

procesos cognitivos que intervienen en el aprendizaje y las estrategias que el docente puede promover para favorecerlos.

Además, los docentes con menor experiencia profesional reportan con mayor frecuencia este tipo de creencias. Esto podría deberse a que, en las primeras etapas de la carrera docente, aún no se han acumulado suficientes experiencias cotidianas en el aula que permitan reconocer la necesidad de enfoques más formativos y flexibles. Asimismo, es posible que la formación inicial no haya brindado oportunidades suficientemente amplias para profundizar en los principios pedagógicos del enfoque formativo.

Lo anterior pone en relieve una necesidad crítica: fortalecer la formación docente a lo largo de toda la trayectoria profesional, no solo en términos de conocimientos técnicos y creencias pedagógicas, sino también a través de procesos que desarrollen la autorreflexión profesional. Esto requiere examinar las condiciones en que se forman los futuros docentes y la coherencia entre las experiencias de evaluación y los enfoques formativos que se espera implementen en el aula. Como muestran algunos estudios, en la formación inicial docente, las prácticas evaluativas no siempre reflejan plenamente los principios del enfoque formativo (Eguren, 2025; Pease, 2018). En esa misma línea, los programas de formación en servicio y los espacios de trabajo colegiado constituyen oportunidades para consolidar enfoques de evaluación más formativos y coherentes con los propósitos del aprendizaje.

En cuanto a la relación entre las creencias docentes y el rendimiento de los estudiantes, se encontró que quienes asisten a escuelas cuyos docentes sostienen concepciones más restrictivas o parciales sobre la evaluación, tienden a evidenciar un menor logro de aprendizajes en Lectura y Matemática. En contextos donde predominan estas ideas, es posible que los estudiantes tengan menos claridad sobre los aprendizajes esperados, reciban menos retroalimentación útil y oportuna, y dispongan de menos oportunidades para demostrar su progreso. Este resultado se alinea con un metaanálisis que muestra que las estrategias más efectivas para mejorar el aprendizaje son aquellas que promueven la autoevaluación del estudiante y una retroalimentación sostenida durante el proceso, y no aquellas centradas únicamente en la verificación del desempeño (Lee et al., 2020).

Según esta evidencia, los resultados del presente análisis muestran que la asociación negativa entre las creencias docentes y los aprendizajes de los estudiantes se mantiene incluso al considerar otros factores del contexto escolar. En particular, el efecto persiste al controlar por el nivel socioeconómico promedio, lo que sugiere que esta relación no se explica únicamente por las diferencias socioeconómicas entre escuelas. Además, el patrón se observa de manera consistente en los distintos estratos de la escuela (público urbano, público rural y privado), lo que refuerza la generalización del resultado.

Estos hallazgos no solo describen una tendencia relevante, sino que también evidencian la necesidad de intervenir sobre los factores que podrían estar limitando la comprensión docente de lo que implica la evaluación con enfoque formativo. En primer lugar, resulta prioritario fortalecer los espacios de formación inicial y continua, de modo que los docentes

puedan consolidar conocimientos, habilidades y disposiciones coherentes con dicho enfoque. Esto requiere no solo incorporar contenidos teóricos, sino también promover experiencias de aprendizaje en las que los futuros docentes vivan prácticas evaluativas alineadas con los principios formativos y desarrollen las competencias necesarias para reflexionar sobre su propia práctica. Los fundamentos pedagógicos del Diseño Curricular Básico Nacional de la Formación Inicial Docente contemplan la evaluación formativa como parte de la preparación de los futuros docentes (Minedu, 2019), lo cual constituye un avance, aunque todavía requiera fortalecerse para consolidar su implementación efectiva.

En segundo lugar, se necesita de acciones orientadas a consolidar un significado compartido de los conceptos vinculados a la evaluación formativa. Esto resulta particularmente importante en la retroalimentación, donde se observó una diversidad de interpretaciones, algunas de las cuales distorsionan o reducen su función formativa. También en la noción de evidencia del progreso, reflejada en prácticas como asignar notas más bajas al inicio con la intención de mostrar luego un avance. Estos ejemplos muestran la necesidad de contar con referentes claros que reduzcan ambigüedades sobre algunos componentes del enfoque formativo. Sin un significado compartido, los mensajes difícilmente se trasladan de manera uniforme al aula, pues tienden a ser reinterpretados según la experiencia y el contexto de cada docente (Ball et al., 2012). En este sentido, el desarrollo de referentes comunes, como rúbricas modelo, guías de retroalimentación o progresiones de aprendizaje, contruidos desde la articulación entre los distintos niveles de gestión (nacional, regional, local y escuelas), podría favorecer la consolidación de un significado compartido.

Finalmente, las dimensiones identificadas en este estudio constituyen un punto de partida para analizar de manera sistemática las creencias docentes sobre la evaluación en el sistema educativo peruano. En adelante, se recomienda explorar aspectos no abordados en este análisis y poner en diálogo estos hallazgos con estudios internacionales sobre creencias similares. Esto permitiría avanzar hacia una comprensión integral y comparada del modo en que los docentes entienden y practican la evaluación.

Para interpretar adecuadamente los resultados, es importante considerar algunas limitaciones del estudio. La principal es que la información se basa en el autoreporte, lo que puede introducir sesgos de deseabilidad social o respuestas socialmente esperadas. Además, las encuestas, por sí solas, no permiten capturar en profundidad las intenciones ni los significados que los docentes otorgan a sus prácticas de evaluación. Otro aspecto a considerar es que, dado el carácter observacional y transversal del estudio, no es posible establecer relaciones de causalidad entre las creencias docentes y el rendimiento de los estudiantes, sino únicamente asociaciones. En este sentido, se debe complementar estos hallazgos con entrevistas, estudios de caso o diseños longitudinales que permitan comprender mejor cómo se configuran las creencias y cómo estas se traducen en prácticas en el aula.

En resumen, los resultados muestran que las creencias docentes acerca de la evaluación no son un aspecto marginal, sino un factor que puede incidir en la calidad de las experiencias de

aprendizaje de los estudiantes. Reconocer y trabajar sobre estas creencias es clave para avanzar hacia un sistema educativo que también apoye el aprendizaje de manera continua y significativa.

Agradecimientos

El autor del estudio agradece a Andrés Burga, Giovanna Moreano, Giovanna Talla, Luz Huanca, Tania Pacheco, Teresa Torres, Víctor Salazar y Yoni Arámbulo por los comentarios y aportes brindados durante la elaboración de esta investigación.

Referencias

- Baird, J.-A., Andrich, D., Hopfenbeck, T. N., y Stobart, G. (2017). Assessment and learning: fields apart? *Assessment in Education: Principles, Policy & Practice*, 24(3), 317-350. <https://doi.org/10.1080/0969594X.2017.1319337>
- Ball, S. J., Maguire, M., y Braun, A. (2012). *How schools do policy: policy enactments in secondary schools*. Routledge.
- Barnes, N., Fives, H., y Dacey, C. (2014). Teachers' beliefs about assessment. En H. Fives y M. G. Gill (Eds.), *International handbook of research on teachers' beliefs* (pp. 284-300). Routledge.
- Barnes, N., Fives, H., y Dacey, C. M. (2017). U.S. teachers' conceptions of the purposes of assessment. *Teaching and Teacher Education*, 65, 107-116. <https://doi.org/10.1016/j.tate.2017.02.017>
- Bates, D., Mächler, M., Bolker, B., y Walker, S. (2025). *lme4: Linear Mixed-Effects Models using 'Eigen' and S4* [R package version 1.1-41]. <https://CRAN.R-project.org/package=lme4>
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5-25. <https://doi.org/10.1080/0969594X.2010.513678>
- Black, P., y Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21, 5-31. <https://doi.org/10.1007/s11092-008-9068-5>
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3-12. <https://doi.org/10.1111/j.1745-3992.2010.00195.x>
- Brown, G. (2004). Teachers' conceptions of assessment: implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301-318. <https://doi.org/10.1080/0969594042000304609>
- Brown, G., Gebril, A., y Michaelides, M. P. (2019). Teachers' conceptions of assessment: a global phenomenon or a global localism. *Frontiers in Education*, 4, Article 16.
- Buehl, M., y Beck, J. (2014). The relationship between teachers' beliefs and teachers' practices. En H. Fives y M. G. Gill (Eds.), *International handbook of research on teachers' beliefs* (pp. 66-84). Routledge.
- Dolin, J., Black, P., Harlen, W., y Tiberghien, A. (2018). Exploring relations between formative and summative Assessment. En J. Dolin y R. Evans (Eds.), *Transforming assessment* (pp. 53-80). Springer International Publishing.
- Eguren, M. (2025). Enseñar a enseñar. *Argumentos*, 5(2), 5-29. <https://doi.org/10.46476/ra.v5i2.224>
- Eguren, M., de Belaunde, C., y González, N. (2019). *Leyendo al Estado desde el aula. Maestros, pedagogía y ciudadanía*. Instituto de Estudios Peruanos.
- Flora, D. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science*, 3(4), 484-501. <https://doi.org/10.1177/2515245920951747>
- González, N., Moscoso, M., y Aragón, J. (2022). *Docentes no titulados: trayectoria educativa, prácticas pedagógicas y alternativas de profesionalización* [Educación, 24]. Instituto de Estudios Peruanos. <https://hdl.handle.net/20.500.14660/1221>
- Hattie, J., y Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan*, 89(2), 140-145. <https://doi.org/10.1177/003172170708900210>
- Hidalgo, N., y Murillo, F. (2021). Assessment conceptions and practices: perspectives of primary school teachers and students. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio En Educación*, 15(1). <https://doi.org/10.15366/reice2017.15.1.007>
- Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241-253. <https://doi.org/10.3102/0013189X20912798>

- Laveault, D., y Allal, L. (2016). Implementing assessment for learning: Theoretical and practical issues. En D. Laveault y L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 1-18). Springer International Publishing. https://doi.org/10.1007/978-3-319-39211-0_1
- Lee, H., Chung, H. Q., Zhang, Y., Abedi, J., y and, M. W. (2020). The effectiveness and features of formative assessment in US K-12 education: A systematic review. *Applied Measurement in Education*, 33(2), 124-140. <https://doi.org/10.1080/08957347.2020.1732383>
- León, E. (2018). *El fenómeno ECE y sus efectos en las prácticas docentes*. En Acción, GRADE, Tarea. http://www.grade.org.pe/wp-content/uploads/FORGE_El-fenomeno-ECE-Eduardo-Leon.pdf
- Llanos, F., y Tapia, J. (2021). *La retroalimentación dialógica: algunos hallazgos en escuelas multigrado* (Informe técnico). Proyecto CREER / GRADE. <https://www.grade.org.pe/creer/archivos/Retroalimentaci%C3%B3n-Llanos-y-Tapia.pdf>
- Looney, A., Cumming, J., van Der Kleij, F., y Harris, K. (2018). Reconceptualising the role of teachers as assessors: teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, 25(5), 442-467. <https://doi.org/10.1080/0969594X.2016.1268090>
- Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., y Nagengast, B. (2009). Doubly-latent models of school contextual effects: integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, 44(6), 764-802. <https://doi.org/10.1080/00273170903333665>
- Ministerio de Educación del Perú. (2014). *Marco de Buen Desempeño Docente: para mejorar tu práctica como maestro y guiar el aprendizaje de tus estudiantes*. <https://hdl.handle.net/20.500.12799/6531>
- Ministerio de Educación del Perú. (2017). *Currículo Nacional de la Educación Básica*. <https://hdl.handle.net/20.500.12799/4551>
- Ministerio de Educación del Perú. (2018). *Desafíos en la medición y el análisis del estatus socioeconómico de los estudiantes peruanos*. Oficina de Medición de la Calidad de los Aprendizajes. <http://umc.minedu.gob.pe/desafios-en-la-medicion-y-el-analisis-del-estatus-socioeconomico-de-los-estudiantes-peruanos/>
- Ministerio de Educación del Perú. (2019). *Diseño Curricular Básico Nacional de la Formación Inicial Docente: Programa de Estudios de Educación Primaria*. <https://hdl.handle.net/20.500.12799/6671>
- Ministerio de Educación del Perú. (2020). *Norma que regula la evaluación de las competencias de los estudiantes de la educación básica*. <https://www.gob.pe/institucion/minedu/normas-legales/5518274-048-2024-minedu>
- Ministerio de Educación del Perú. (2024). *Reporte de resultados del Monitoreo de Prácticas Escolares 2023*. Oficina de Seguimiento y Evaluación Estratégica. <https://hdl.handle.net/20.500.12799/10541>
- Ministerio de Educación del Perú. (2025). *Reporte técnico de la Evaluación Nacional de Logros de Aprendizaje de Estudiantes (ENLA) 2024*. Oficina de Medición de la Calidad de los Aprendizajes. http://umc.minedu.gob.pe/wp-content/uploads/2025/05/RT_ENLA2024.pdf
- Monteiro, V., Mata, L., y Santos, N. N. (2021). Assessment conceptions and practices: perspectives of primary school teachers and students. *Frontiers in Education*, 6, Article 631185. <https://doi.org/10.3389/feduc.2021.631185>
- Näsström, G., Andersson, C., Granberg, C., Palm, T., y Palmberg, B. (2021). Changes in student motivation and teacher decision making when implementing a formative assessment practice. *Frontiers in Education*, 6, Article 616216. <https://doi.org/10.3389/feduc.2021.616216>
- Pajares, M. F. (1992). Teachers' beliefs and educational research: cleaning up a messy construct. *Review of Educational Research*, 62(3), 307-332. <https://doi.org/10.3102/00346543062003307>
- Pease, M. E. (2018). *El enfoque formativo en la formación inicial docente de una facultad de educación de una universidad privada de Lima* [Tesis de Maestría, Pontificia Universidad Católica del Perú]. Repositorio Digital de Tesis y Trabajos de Investigación PUCP. <http://hdl.handle.net/20.500.12404/12861>
- R Core Team. (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>

- Ravela, P. (2009). Consignas, devoluciones y calificaciones: los problemas de la evaluación en las aulas de educación primaria en América Latina. *Páginas de Educación*, 2(1), 49-89. <https://doi.org/10.35362/rie7623072>
- Ravela, P., Leymoníé, J., Viñas, J., y Haretche, C. (2014). La evaluación en las aulas de secundaria básica en cuatro países de América Latina. *Propuesta Educativa*, (41), 20-45. <https://propuestaeducativa.flacso.org.ar/wp-content/uploads/2019/12/41-dossier-ravelayotros.pdf>
- Ravela, P., Picaroni, B., y Loureiro, G. (2017). *¿Cómo mejorar la evaluación en el aula? Reflexiones y propuestas para trabajo de docentes*. Grupo MAGRO Editores.
- Remesal, A. (2011). Primary and secondary teachers' conceptions of assessment: a qualitative study. *Teaching and Teacher Education*, 27(2), 472-482. <https://doi.org/10.1016/j.tate.2010.09.017>
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36. <https://doi.org/10.18637/jss.v048.i02>
- Schellekens, L. H., Bok, H. G., de Jong, L. H., van der Schaaf, M. F., Kremer, W. D., y van der Vleuten, C. P. (2021). A scoping review on the notions of Assessment as Learning (AaL), Assessment for Learning (AfL), and Assessment of Learning (AoL). *Studies in Educational Evaluation*, 71, Article 101094. <https://doi.org/10.1016/j.stueduc.2021.101094>
- Xu, Y., y Brown, G. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-162. <https://doi.org/10.1016/j.tate.2016.05.010>
- Yan, Z., y Chiu, M. M. (2023). The relationship between formative assessment and reading achievement: A multilevel analysis of students in 19 countries/regions. *British Educational Research Journal*, 49(1), 186-208. <https://doi.org/10.1002/berj.3837>
- Yan, Z., Li, Z., Panadero, E., Yang, M., Yang, L., y Lao, H. (2021). A systematic review on factors influencing teachers' intentions and implementations regarding formative assessment. *Assessment in Education: Principles, Policy & Practice*, 28(3), 228-260. <https://doi.org/10.1080/0969594X.2021.1884042>

Anexos

Tabla A.1

Resultados completos de las regresiones entre las creencias docentes sobre la evaluación y el rendimiento en Lectura y Matemática

	Lectura				Matemática			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Intercepto	475,02 (0,35)	475,51 (0,34)	468,36 (0,80)	461,75 (0,76)	452,18 (0,45)	452,92 (0,45)	440,63 (1,07)	430,04 (1,10)
Creencias docentes		-7,52 (0,38)	-7,13 (0,38)	-5,54 (0,27)		-11,19 (0,49)	-10,53 (0,50)	-8,09 (0,40)
Sexo docente (Mujer)			13,25 (0,35)	7,93 (0,33)			14,92 (0,47)	8,81 (0,45)
Título pedagógico (Sí)			-8,03 (0,70)	2,20 (0,61)			-6,46 (0,94)	3,84 (0,87)
Años de experiencia (6 a 20)			4,31 (0,54)	4,32 (0,48)			7,14 (0,72)	6,86 (0,67)
Años de experiencia (>20)			7,01 (0,55)	7,13 (0,50)			10,78 (0,73)	10,40 (0,69)
Escuela pública rural				7,78 (1,04)				14,75 (1,56)
Escuela pública urbana				9,84 (0,75)				19,07 (1,13)
ISEP				40,54 (0,49)				47,21 (0,73)
Varianza residual	4676	4676	4672	4675	8343	8343	8339	8341
Varianza entre escuelas	1813	1764	1630	602	3060	2952	2813	1511
AIC	5869070	5868683	5523599	5504260	6164580	6164075	5802271	5786943
BIC	5869103	5868728	5523688	5504382	6164614	6164120	5802359	5787065

Tabla A.2

Resultados completos de las regresiones entre las creencias docentes sobre la evaluación y el rendimiento, según estrato de la escuela

	Lectura			Matemática		
	Privada	Pública rural	Pública urbana	Privada	Pública rural	Pública urbana
Intercepto	476,90 (1,14)	461,36 (2,72)	469,51 (1,33)	452,35 (1,59)	432,13 (3,95)	446,57 (1,84)
Creencias docentes	-5,36 (0,43)	-5,56 (0,46)	-5,65 (0,51)	-6,87 (0,61)	-8,82 (0,69)	-8,05 (0,81)
Sexo docente (Mujer)	-0,64 (0,94)	9,74 (0,83)	8,62 (0,39)	-3,17 (1,31)	12,69 (1,19)	9,39 (0,53)
Título pedagógico (Sí)	3,10 (0,71)	12,69 (2,32)	1,45 (1,17)	6,23 (1,00)	21,06 (3,34)	1,32 (1,59)
Años de experiencia (6 a 20)	0,78 (0,75)	7,10 (1,40)	6,95 (0,68)	-0,84 (1,04)	10,17 (2,02)	11,89 (0,92)
Años de experiencia (>20)	2,45 (0,99)	9,89 (1,47)	9,64 (0,67)	-0,54 (1,37)	14,24 (2,12)	15,27 (0,91)
ISEP	33,79 (0,76)	46,66 (0,95)	41,53 (0,79)	38,13 (1,08)	57,33 (1,41)	46,78 (1,28)
Varianza residual	4582	4537	4743	8073	8349	8442
Varianza entre escuelas	432	935	498	926	2239	1400
AIC	1398935	895312	3209189	1468625	944424	3372816
BIC	1399023	895396	3209284	1468713	944507	3372911