

Diseño y validación de un cuestionario para evaluar desempeño docente asociado a las prácticas evaluativas formativas

Claudio Cerón Urzua¹, Marco Cossio-Bolaños², Paz Pezoa-Fuentes³, Rossana Gómez-Campos⁴

Recibido: Septiembre 2019 / Evaluado: Enero 2020 / Aceptado: Febrero 2020

Resumen. La evaluación formativa es una práctica evaluativa áulica para el mejoramiento del aprendizaje. El objetivo de este trabajo es reportar el proceso de validación de un cuestionario de evaluación formativa con base en la opinión de los estudiantes sobre los desempeños del profesorado. El método con el que se trabajó es de carácter cuantitativo, (exploratorio) en base a un Análisis Factorial Confirmatorio a través de modelos de ecuaciones estructurales (SEM). Se trabajó con una muestra de 501 estudiantes de 5° a 8° básico (10 a 13 años) de la comuna de Yerbos Buenas en Chile. Los resultados en torno a la construcción de un cuestionario de evaluación formativa a nivel de aula, aportan con evidencia de validez de constructo para el instrumento, lo que hace de este cuestionario un instrumento válido (CIF 0,967; RMR 0,04; RMSEA 0,041) y fiable (Alfa de cronbach: 0,93). Las evidencias generadas de la investigación confirman que existen seis dimensiones útiles para evaluar el desempeño docente relacionadas con la práctica evaluativa formativa áulica y que con ellas se puede determinar el nivel de desempeño del profesorado. Los resultados sugieren su uso y aplicación en muestras con similares características.

Palabras claves: Validación, Análisis Factorial Confirmatorio, Ecuaciones Estructurales, Evaluación Formativa, Desempeño Docente

[en] Design and validation of a questionnaire to evaluate teaching performance associated with training practices

Abstract. Formative evaluation is an evaluative practice for the improvement of learning. The objective of this work is to report the validation process of a formative evaluation questionnaire based on the opinion of the students about the teachers' performances. The method used was quantitative (exploratory), based on a Confirmatory Factor Analysis through structural equation models (SEM). We worked with a sample of 501 students from 5th to 8th grade (10 to 13 years old) from the Yerbos Buenas district in Chile. The results around the construction of a formative evaluation questionnaire at the classroom level provide evidence of construct validity for the instrument, which makes this questionnaire a valid instrument (CIF 0.967; RMR 0.04; RMSEA 0.041) and reliable (Cronbach's alpha: 0.93). The evidence generated from the research confirms that there are six useful dimensions to evaluate teaching performance related to formative evaluative practice in the classroom and that the level of performance of teachers can be determined with them. The results suggest its use and application in samples with similar characteristics.

Keywords: Validation; Confirmatory Factor Analysis; Structural Equations; Formative Evaluation; Teaching Performance

Sumario. 1. Introducción. 2. Metodología. 2.1. Tipo de estudio y muestra. 2.2. Técnicas instrumentos. 2.3. Análisis de datos. 3. Resultados. 4. Discusión y conclusión. 5. Referencias bibliográficas

Cómo citar: Cerón Urzua, C., Cossio-Bolaños, M., Pezoa-Fuentes, P., Gómez-Campos, R. (2020). Diseño y validación de un cuestionario para evaluar desempeño docente asociado a las prácticas evaluativas formativas. *Revista Complutense de Educación*, 31(4), 463-472.

1. Introducción

Las competencias del profesorado están enmarcadas dentro de la escuela y se conjugan dentro de múltiples situaciones pedagógicas vinculadas al diseño de la enseñanza, la mediación de los aprendizajes y la evaluación, entre otras. En

¹ Universidad Católica del Maule (Chile).
Email: cceronu@ucm.cl

² Universidad Católica del Maule (Chile).
Email: mcossio1972@hotmail.com

³ Universidad Católica del Maule (Chile).
Email: ppezoa@ucm.cl

⁴ Universidad Católica del Maule (Chile).
Email: rossaunicamp@gmail.com

general, el profesor es un docente que se desempeña en un campo pedagógico de inter-relaciones, lo que permite definir el desempeño del docente y ende comprender que existe una relación dialéctica entre el sujeto (docente), el objeto como aprendizaje mediado y los estudiantes que aprenden.

En ese sentido se define como desempeño docente el saber “actuar eficazmente” (Perrenoud, 1998) dentro de un contexto, en este caso de un ambiente áulico en contexto escolar. Esto implica el desarrollo de desempeños pedagógicos áulicos asociadas a un aspecto específico definido como proceso de enseñanza-aprendizaje, dentro de estos desempeños podemos mencionar algunos como, gestionar actividades para mediar los procesos de aprendizaje y evaluar formativamente los resultados de aprendizaje a partir de acciones reflexivas y de retroalimentación.

Autores como Perrenoud (2004) consideran que una de las diez competencias pedagógicas más relevantes está asociada a “gestionar la progresión de los aprendizajes” y que tiene relación con generar condiciones propicias para aprender. Dentro de esta competencia destaca un desempeño vinculado con la evaluación.

La evaluación es una acción conceptualmente única que, dependiendo del momento o de la necesidad, adopta funciones diversas (Lukas & Santiago, 2009) adquiriendo importancia la temporalidad, sus actores y su finalidad. Dentro de este último punto, la evaluación se divide en diagnóstica, formativa y Sumativa.

La evaluación diagnóstica permite evaluar aspectos de entrada como los inicios de un curso, el diseño de enseñanza, las concepciones previas, entre otros. En tanto la evaluación sumativa tiene como propósito determinar niveles de rendimiento y hace referencia a un juicio en torno a un proceso que ha finalizado, por lo que es retrospectivo (Núñez & Quiñones, 2012).

Por su parte la evaluación formativa tiene un propósito distinto. El concepto de evaluación formativa, se ha de-construido con el correr del tiempo, paso de ser durante la década de los años 60 del siglo XX, un tipo de evaluación que intentaba buscar información sobre el proceso, oponiéndose a la evaluación que sirve para valorar el resultado final, hasta su posterior ampliación conceptual, durante la década del 90, como una instancia en el que el docente puede tomar decisiones instruccionales a partir de la información que obtiene del proceso (Brookhar, 2009).

Shepard, Hammerness, Darling-Hammond y Rust (2005) desde el inicio definieron a la evaluación formativa como una evaluación que es llevada a cabo durante el proceso instructivo con el propósito de mejorar la enseñanza o el aprendizaje, lo que significó un avance sustancial desde el punto de vista epistemológico del concepto, ya que permitió establecer una finalidad y constituirse como una alternativa al enfoque por medición.

Autores como Black y Wiliam (1998) y Brookhart (2009), han consolidado la idea de que la evaluación formativa no sólo es efectiva en la obtención y comunicación de información cognitiva, sino que además es el motor motivador en el aprendizaje, ya que, desprendida de su valor punitivo, la calificación, la evaluación formativa es capaz de impactar afectivamente a los estudiantes para que sigan aprendiendo.

Por otro lado, la evaluación formativa permite que los docentes hagan ajustes a la instrucción a partir de la evidencia recolectada, entregándoles a los estudiantes retroalimentación que les permita mejorar sus aprendizajes. Para Black y Wiliam (2009) la evaluación formativa permite recolectar evidencia acerca del logro en el aprendizaje y que, al ser interpretada y utilizada (por docentes y estudiantes), ayuda para tomar decisiones sobre los próximos pasos a seguir en el proceso de enseñanza y aprendizaje. Aquí el uso de la evidencia recolectada a través de la evaluación no sólo sirve para informar, sino también transformar la práctica pedagógica y mejorar el desempeño de los estudiantes.

En palabras de Perrenoud “una de las competencias cruciales, en la enseñanza, es saber regular los procesos de aprendizaje más que ayudar al éxito de la actividad” (2004, p. 41). Podemos decir entonces que, la evaluación formativa se sitúa como un desempeño docente práctico y cercano a la singularización del individuo que aprende, es decir, no estandarizada y más que promover una medición tiene como propósito regular los procesos.

En Chile, existen ciertas problemáticas que deben resolverse. Según UNICEF (2016) se debe revisar el sistema de evaluación ya que está orientado a la calificación, principalmente de resultados y productos finales. Sabiendo esto, cabe preguntarse ¿Cómo se están desempeñando los docentes en nuestro país, en el ámbito de la evaluación formativa?

Los informes de evaluación docente que realiza el Estado en Chile, demuestran qué aspectos relacionados con la evaluación para el aprendizaje son una debilidad en las escuelas, declarando que el desempeño vinculado a la evaluación formativa definido como: “Reflexión a partir de los resultados de la evaluación” es el más descendido. (Manzi, González & Sun, 2011; Herrada, 2015).

Esta dimensión está directamente vinculada con la responsabilidad docente frente a los resultados de los alumnos y la retroalimentación efectuada a sus estudiantes. En este sentido, destacamos de manera particular que el enfoque pedagógico de la evaluación está vinculada directamente a la evaluación formativa y a la retroalimentación de aprendizajes a nivel de aula. Otro estudio publicado por Ravela, Leymonié, Viñas y Haretche (2014) en torno a la evaluación de aula en países como Chile, Perú, Colombia y Uruguay, demostró que en Chile prevalece un enfoque por medición, cuyo énfasis está dado en la calificación y el uso de instrumentos asociados al ítem de preguntas cerradas.

Dentro de la escuela, la evaluación docente debe ser una prioridad, con el propósito de desarrollar una cultura evaluativa y reflexiva del quehacer docente (Rueda, Luna, García & Loredó, 2010), donde lo fundamental sea la valorización y reflexión de las prácticas pedagógicas que orientan el mejoramiento de la calidad educativa (Carlos-Guzmán, 2016). A partir de lo anterior es relevante evaluar el desempeño de los docentes puesto que los desempeños a nivel de aula permiten un mayor aseguramiento de la calidad de los aprendizajes y por ende una mejor gestión por parte de los equipos de gestión de sus docentes en torno al desarrollo del currículo y las prácticas pedagógicas para una mejor toma de decisiones promover la regulación y autorregulación docente.

En consecuencia, contar con un instrumento de evaluación del desempeño docente en torno a la evaluación formativa en la medida que el instrumento responde a un modelo teórico y cuenta con la validez en los criterios psicométricos, tomando la percepción de los estudiantes tiene múltiples ventajas ya que permite valorar la práctica docente (García & Medécigo, 2014). Sin embargo, esto debe ir acompañado de orientaciones y apoyo técnico hacia el docente que le permita identificar dónde se encuentra y qué es lo que debe lograr.

Desde esta perspectiva, en Chile prevalece un enfoque por medición sustentado por una fuerte exigencia curricular, en donde predominan los indicadores de evaluación y el avance del desarrollo escolar estandarizado. De hecho, no se considera la singularidad del alumno, por lo que se establece una brecha respecto de lo que los y las estudiantes saben y pueden hacer y acercándose a lo que debiesen saber en vinculación con el progreso de un curso estándar (Equipo D. T., 2015).

Dentro de este contexto, comprendemos la importancia de la evaluación formativa y la problemática que se establece en Chile, al coexistir una alta necesidad por mejorar los aprendizajes, pero a la vez, una cultura escolar instalada sobre la calificación y la educación estandarizada. A partir de ello, es importante contar con un instrumento que mida este desempeño para poder identificar qué dimensiones son las que están asociadas a la práctica evaluativa de carácter formativo y cuáles son los aspectos que deben mejorar.

En ese contexto, de acuerdo, a los informes de la evaluación docente que realiza el Estado, que los docentes no se desempeñan competentemente en la gestión los procesos reflexivos asociados a los resultados de aprendizaje de los estudiantes y que por ende sobresale el enfoque por medición. Sin embargo, desconocemos cuales son todas las dimensiones relacionadas con esta mala práctica.

Sostenemos como hipótesis que existen variadas dimensiones que contribuyen al buen desempeño de un docente y que están relacionadas con: la comunicación de los resultados de las pruebas de rendimiento a través de la retroalimentación; la anticipación de las dificultades en el aprendizaje de manera pro-activa; la evaluación formativa de proceso de manera interactiva vinculada a la construcción de aprendizajes autónomos; la evaluación formativa metacognitiva, asociada al proceso interno de la información; la evaluación formativa retroactiva, que retroalimenta los resultados o producto durante el proceso de aprendizaje durante la clase; y, la evaluación formativa ajustada a las necesidades de los estudiantes y su trayectoria escolar.

Por lo tanto, ante la necesidad y carencia de instrumentos que permitan evaluar el desempeño docente durante las prácticas profesionales, este estudio se propuso como objetivo, proponer un instrumento que permita evaluar las prácticas del profesorado en aspectos evaluativos, así como de retroalimentación de aprendizajes es crucial para la toma decisiones educativas.

2. Metodología

2.1. Tipo de estudio y muestra

Se efectuó un estudio cuantitativo de tipo descriptivo exploratorio. El universo estuvo conformado por 1800 escolares. La selección de la muestra fue de tipo probabilístico (sistemático) teniendo un IC 95%. Se seleccionó a 501 alumnos (272 hombres y 229 mujeres) de escuelas rurales del distrito de Longaví (Talca, Chile). Estos corresponden al 27.8% del universo total. Esta región está localizada en la séptima región (Maule) y su principal actividad es la agricultura y el turismo.

Los estudiantes presentan un rango de edad de 10 a 13 años y están cursando el 5°, 6°, 7° y 8° básico de Colegios Municipales. Estos alumnos se caracterizan por pertenecer a la condición socioeconómica media baja, puesto viven en zonas rurales de la región.

2.2. Técnicas instrumentos

Se utilizó la técnica de la encuesta para medir la variable desempeño docente en las prácticas evaluativas. Se elaboró un instrumento (cuestionario) basado en la teoría desarrollada por varios autores (Brookhart, 2007; Hattie & Timperley, 2007; Shepard, 2007; Jorba & Sanmartí, 2008; Black & Wiliam, 1998, 2009; Wiliam, 2009). A partir de dicha información se operacionalizó en seis dimensiones y sus respectivos indicadores. La tabla 1, muestra las dimensiones y sus indicadores.

Tabla 1. Dimensiones relacionadas con la evaluación formativa

Dimensión	Ítem
Evaluación formativa asociada a la calificación.	1. Después de la aplicación de una evaluación calificada, el profesor(a) me llama para retroalimentarme personalmente.
	2. Cuando me entregan una prueba, el profesor(a) me escribe una breve nota en la misma prueba explicando mis principales errores y dificultades.
	3. Después de la aplicación de una evaluación calificada, el profesor(a) retroalimenta de manera verbal a todo el curso, explicando los principales errores de la prueba.
Evaluación formativa proactiva	4. Antes de trabajar en una actividad, me entregan las instrucciones escritas con lo que debes lograr en la tarea
	5. Antes de trabajar en una actividad, el profesor(a) te explica las instrucciones y tienes la oportunidad de preguntar
	6. Cuando me evalúan con una pauta de observación, el profesor(a) la comparte antes de aplicarla lo que me permite leerla y poder consultar lo que no comprenda.
Evaluación formativa interactiva	7. Cuando estás trabajando en una actividad, el profesor(a) te va monitoreando (revisa tu trabajo) para retroalimentarte de inmediato.
	8. Cuando le comunicas al docente que no estás comprendiendo, este te explica con otras palabras o te ofrece ejemplos para que puedas comprender
	9. Cuando un(a) estudiante se equivoca (comete un error), el profesor(a) lo expone como un problema e involucra al curso para analizarlo y responderlo entre todos.
	10. Cuando el grupo tiene problemas en su tarea, el profesor(a) le sugiere una estrategia para que el grupo corrija su respuesta y pueda avanzar.
Evaluación formativa metacognitiva	11. Cuando el profesor(a) evalúa formativamente (sin nota) a través preguntas de respuestas cortas o ejercicios breves se preocupa de la razón o el porqué de mi respuesta que de corregir el error
	12. Cuando el profesor(a) evalúa formativamente a través preguntas de verdadero y falso o selección múltiple se preocupa de la razón o el porqué de mi respuesta que de corregir el error
	13. Cuando el profesor(a) evalúa formativamente a través de preguntas abiertas o resolución de problemas, se preocupa de la razón o el porqué de mi respuesta que de corregir el error
	14. Cuando el profesor(a) evalúa formativamente productos como informes o cartas, se preocupa de evaluar tanto el proceso (el paso a paso) como el resultado final.
Evaluación formativa retroactiva	15. Cuando respondes incorrectamente, el profesor(a) te retroalimenta los errores.
	16. Cuando respondes incorrectamente, el profesor(a) te entrega información para que tomes conciencia del error, comprendiendo por qué esta malo o te equivocaste.
	17. Cuando respondes incorrectamente, el profesor(a) te ayuda entregándote una estrategia (procedimientos) para poder resolver el error, sin dar la respuesta
	18. Cuando respondes incorrectamente, el profesor(a) te ayuda a buscar la respuesta correcta por medio de otras preguntas que te guíen a la solución.
Evaluación formativa ajustada	19. Cuando el profesor(a) retroalimenta observas que lo hace de manera distinta para cada estudiante de acuerdo a su nivel y necesidad, nunca igual para todos.
	20. Cuando el profesor(a) te retroalimenta y aun así no comprendes, el profesor(a) cambia su explicación o estrategia y te entrega nuevos ejemplos.
	21. Cuando algunos de tus compañeros no puede realizar la tarea por completo el profesor(a) acomoda o adapta la actividad o tiempos para que él pueda realizarla

El instrumento evidenció 21 preguntas. Las alternativas fueron escala tipo Likert: Donde 1 indica Nunca, 2 Rara Vez, 3 Ocasionalmente, 4 Frecuentemente y 5 Muy Frecuentemente. Este instrumento tiene por objetivo determinar el desempeño del profesorado en relación a la evaluación formativa. El cuestionario fue aplicado a estudiantes en sus respectivos colegios (horario de clases). La técnica utilizada fue a lápiz y papel. Un encuestador estuvo a cargo de todo el proceso, en el que explicó los procedimientos y despejó dudas. Se les dio 10 a 15 minutos para responder dicho instrumento. El procedimiento para validar fue la validez de constructo (análisis factorial confirmatorio). La confiabilidad se determinó por medio de Consistencia interna.

Se incluyeron en el estudio a los escolares que asistieron el día de la evaluación y a los que se encontraban en el rango de edad establecido. Se excluyeron a los escolares que no completaron los cuestionarios. Todo el protocolo se llevó a cabo de acuerdo a Helsinki (1991) para la investigación en seres humanos y el comité de ética de la UCM-Talca.

2.3. Análisis de datos

La prueba de Kolmoronov Smirnov fue utilizada para verificar la normalidad de los datos. Como los datos presentaron una distribución normal ($p > 0,05$), se efectuó el análisis estadístico descriptivo de media aritmética, desviación estándar y rango. Para verificar las diferencias entre cursos, se utilizó Anova de una vía y la prueba

de especificidad de Tukey. El cálculo del modelo de ajuste entregó información de los siguientes valores: Raíz cuadrada media residual (RMR), Error cuadrático medio de aproximación (RMSEA) y el Índice de ajuste comparativo (CIF). Se calculó la fiabilidad por medio de consistencia interna (Alpha de Cronbach). Se consideró significativo en todos los casos ($p < 0.05$). Los cálculos se efectuaron mediante los programas AMOS V.26. y SPSS V 25.

3. Resultados

La tabla 2 muestra los resultados de las dimensiones del instrumento propuesto. Los valores medios y desvíos están organizados según curso y sexo. Se determinó diferencias significativas entre los valores medios del 5° básico con los alumnos de 8° básico ($p < 0.05$). Los alumnos del 5to Básico reflejan valores más altos en relación a los de cursos superiores en alumnos de ambos sexos.

Tabla 2. Características descriptivas de la muestra estudiada.

Dimensión	Hombres (n= 272)															
	5° Básico				6° Básico				7° Básico				8° Básico			
	X	DE	Mín	Máx	X	DE	Mín	Máx	X	DE	Mín	Máx	X	DE	Mín	Máx
Sumativa	10,3 ^{ad}	3	3	15	8,05	2,9	3	15	9,07	2,6	3	15	8,46	2,9	3	14
Proactiva	12,8 ^c	2,3	7	15	12,2	2	8	15	11,8	2,5	5	15	11,2	2,6	5	15
Interactiva	16,9 ^{ac}	3,1	9	20	14,9	3,6	6	20	15	3,1	8	20	14,1	3,7	6	19
Metacognición	15,9 ^{abc}	3,5	7	20	13,8	4,1	4	20	13,4	3,1	7	20	12,6	3,2	5	19
Retroactiva	16,7 ^{abc}	2,8	7	20	14,4	3,9	4	20	14,4	3,6	4	20	13,9	3,4	5	19
Ajustada	12,3 ^{abc}	2,5	5	15	10,9	3,1	4	15	10,6	2,5	4	15	10,1	2,7	3	15
	Mujeres (n= 229)															
Sumativa	12,0 ^{abc}	2,8	5	15	8,52	2,8	4	15	7,7	3	3	14	8,45	2,8	3	14
Proactiva	13,7	1,8	8	15	12,0	2,7	5	15	12,0	2,6	3	15	11,2	2,8	6	15
Interactiva	17,7 ^{abc}	3,3	7	20	15,3	3,1	9	20	15,3	3,4	7	20	14,0	3,6	7	19
Metacognición	15,7 ^{bc}	4,7	4	20	14,2	4,3	4	20	13,6	3,5	6	20	13,3	3,6	7	20
Retroactiva	17,9 ^{abc}	2,9	6	20	14,4	4	5	20	14,7	4,4	5	20	13,6	3,4	6	20
Ajustada	12,4 ^{abc}	2,5	5	15	10,8	2,9	4	15	10,4	2,8	3	15	9,68	2,4	3	15
	Ambos sexos (n = 501)															
Sumativa	11,0 ^{abc}	3	3	15	8,31	2,8	3	15	8,5	2,9	3	15	8,46	2,8	3	14
Proactiva	13,2 ^{abc}	2,2	7	15	12,1	2,4	5	15	11,9	2,5	3	15	11,2	2,7	5	15
Interactiva	17,2 ^{abc}	3,2	7	20	15,1	3,3	6	20	15,1	3,2	7	20	14,1	3,6	6	19
Metacognición	15,8 ^{abc}	4	4	20	14	4,2	4	20	13,5	3,3	6	20	12,9	3,4	5	20
Retroactiva	17,2 ^{abc}	2,9	6	20	14,4	3,9	4	20	14,5	3,9	4	20	13,8	3,4	5	20
Ajustada	12,4 ^{abc}	2,5	5	15	10,8	3	4	15	10,5	2,7	3	15	9,93	2,5	3	15

Leyenda: X: Promedio, DE: Desviación estándar, a: diferencia significativa en relación a 6°, b: diferencia en relación a 7° y c: diferencia en relación a 8° ($P < 0.05$).

En cuanto a los valores del modelo confirmatorio, los resultados por dimensión y grupo de edad del instrumento propuesto, se muestran en la tabla 3. Los valores del CIF son inferiores a 0.96 en ambos modelos desarrollados. La raíz cuadrada media residual (RMR) presenta en ambos modelos valores inferiores a 0.05, y el Error cuadrático medio de aproximación (RMSEA) mostró un valor inferior 0.05.

Tabla 3. Valores del modelo confirmatorio según edad y dimensión

Muestra		α	CCI	GFI	NFI	TLI	CFI	RMSEA	χ^2/gl
Todos		0,93	0,8	0,99	0,98	0,98	0,951	0,049	1,97
Sexo	Mujeres	0,91	0,87	0,99	0,99	0,98	0,96	0,039	1,33
	Hombres	0,93	0,7	0,99	0,97	0,99	0,95	0,032	1,2
Edad (años)	10	0,91	0,77	0,99	0,96	0,99	0,94	0,032	1,18
	11	0,91	0,84	0,98	0,98	0,98	0,942	0,034	1,96
	12	0,91	0,88	0,99	0,98	0,99	0,951	0,034	1,78
	13	0,93	0,79	0,98	0,97	0,99	0,94	0,032	1,56

Nota: CCI = Coeficiente de Correlación Intraclase; GFI = Índice de Bondad de Ajuste; NFI = Índice de Ajuste Normado; TLI = Índice Tucker-Lewis; CFI = Índice de Ajuste Comparativo; RMSEA = Error de Aproximación Cuadrático Medio; χ^2/gl = Chi-cuadrado Normada; α = Coeficiente Alfa de Cronbach

En la figura 1 se muestra los valores que el modelo de AFC. Se evidencia que las saturaciones oscilan entre 0.49 a 0.78 en las seis dimensiones. El test de adecuación de la muestra de Kaiser-Meyer-Olkin fue de 0.954, $X^2= 4262.295$ y el test de esfericidad de Bartlett fue significativo ($p<0.05$). El % de explicación de la varianza del instrumento fue de 60%.

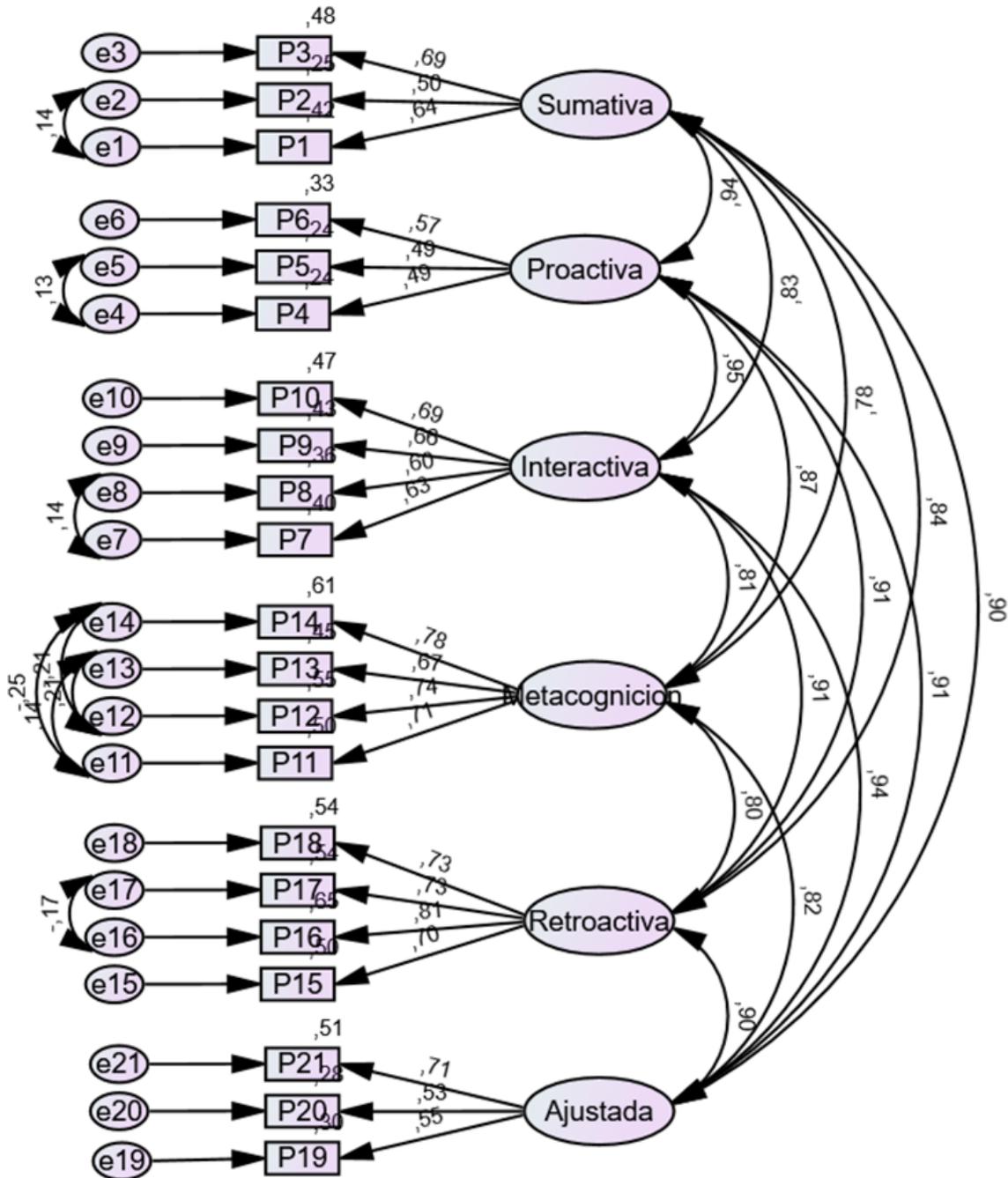


Figura 1. Modelo de ecuaciones estructurales del instrumento elaborado.

En cuanto a la confiabilidad, los valores de Alpha de Cronbach por edad y por cada dimensión que se muestran en la tabla 4 reflejaron valores entre 0.91 a 0.92. Además, el instrumento en su totalidad reflejó un $r= 0.93$. Estos valores reflejan elevados valores de fiabilidad.

Tabla 4. Valores de Alfa de Cronbach del instrumento elaborado

	10 años					11 años					12 años					13 años					Total				
	Media de escala si se elimina el elemento	Varianza de escala si se elimina el elemento	Correlación elemento total	Alfa de Cronbach si el elemento se elimina	Media de escala si se elimina el elemento	Varianza de escala si se elimina el elemento	Correlación elemento total	Alfa de Cronbach si el elemento se elimina	Media de escala si se elimina el elemento	Varianza de escala si se elimina el elemento	Correlación elemento total	Alfa de Cronbach si el elemento se elimina	Media de escala si se elimina el elemento	Varianza de escala si se elimina el elemento	Correlación elemento total	Alfa de Cronbach si el elemento se elimina	Media de escala si se elimina el elemento	Varianza de escala si se elimina el elemento	Correlación elemento total	Alfa de Cronbach si el elemento se elimina	Media de escala si se elimina el elemento	Varianza de escala si se elimina el elemento	Correlación elemento total	Alfa de Cronbach si el elemento se elimina	
P1	83,13	184,35	0,49	0,91	72,33	215,58	0,56	0,90	71,35	194,17	0,48	0,91	67,92	219,41	0,58	0,93	74,00	233,15	0,58	0,92	74,00	233,15	0,58	0,92	
P2	83,44	183,57	0,44	0,91	72,58	219,57	0,39	0,90	71,85	198,50	0,31	0,91	67,96	225,31	0,41	0,93	74,29	236,82	0,44	0,92	74,29	236,82	0,44	0,92	
P3	82,52	188,14	0,53	0,91	71,28	217,03	0,56	0,90	70,42	187,78	0,61	0,90	66,71	214,87	0,66	0,93	73,06	234,35	0,60	0,92	73,06	234,35	0,60	0,92	
P4	82,42	184,85	0,59	0,91	70,78	224,60	0,31	0,91	70,11	198,00	0,37	0,91	66,83	219,94	0,55	0,93	72,84	239,41	0,47	0,92	72,84	239,41	0,47	0,92	
P5	81,92	198,77	0,36	0,91	70,56	228,88	0,29	0,91	69,75	197,25	0,44	0,91	66,27	221,63	0,60	0,93	72,44	244,02	0,46	0,92	72,44	244,02	0,46	0,92	
P6	82,58	183,69	0,52	0,91	71,09	213,21	0,54	0,90	70,38	195,03	0,45	0,91	66,74	216,64	0,59	0,93	73,02	234,85	0,53	0,92	73,02	234,85	0,53	0,92	
P7	82,43	184,09	0,73	0,90	71,23	217,77	0,45	0,90	70,29	196,75	0,44	0,91	66,90	216,41	0,64	0,93	73,03	235,43	0,58	0,92	73,03	235,43	0,58	0,92	
P8	82,13	192,97	0,47	0,91	70,60	228,43	0,32	0,91	69,88	190,79	0,63	0,90	66,60	215,21	0,67	0,93	72,60	239,46	0,55	0,92	72,60	239,46	0,55	0,92	
P9	82,44	182,91	0,72	0,90	71,09	219,21	0,48	0,90	70,47	191,35	0,57	0,90	66,97	219,48	0,55	0,93	73,05	234,52	0,61	0,92	73,05	234,52	0,61	0,92	
P10	82,55	182,84	0,67	0,90	71,29	216,12	0,55	0,90	70,40	193,16	0,55	0,90	66,87	217,85	0,68	0,93	73,10	234,47	0,63	0,92	73,10	234,47	0,63	0,92	
P11	82,69	186,63	0,53	0,91	71,40	211,47	0,62	0,90	70,77	198,10	0,40	0,91	67,28	217,11	0,65	0,93	73,34	234,33	0,59	0,92	73,34	234,33	0,59	0,92	
P12	82,86	181,07	0,59	0,91	71,46	211,26	0,65	0,90	70,81	193,69	0,55	0,90	67,27	220,91	0,58	0,93	73,41	233,18	0,62	0,92	73,41	233,18	0,62	0,92	
P13	82,85	185,73	0,50	0,91	71,39	210,35	0,65	0,90	70,84	193,47	0,60	0,90	67,29	222,41	0,59	0,93	73,40	234,28	0,61	0,92	73,40	234,28	0,61	0,92	
P14	82,53	181,40	0,72	0,90	71,06	215,10	0,61	0,90	70,24	188,00	0,68	0,90	66,65	217,36	0,68	0,93	72,94	233,29	0,67	0,92	72,94	233,29	0,67	0,92	
P15	82,29	189,70	0,52	0,91	71,32	213,15	0,58	0,90	70,61	188,50	0,65	0,90	67,02	217,74	0,63	0,93	73,12	232,24	0,64	0,92	73,12	232,24	0,64	0,92	
P16	82,51	183,54	0,73	0,90	71,25	211,25	0,69	0,90	70,49	187,38	0,70	0,90	67,02	217,70	0,67	0,93	73,13	231,09	0,72	0,92	73,13	231,09	0,72	0,92	
P17	82,38	187,06	0,60	0,91	71,21	211,92	0,65	0,90	70,13	189,32	0,68	0,90	66,80	220,71	0,61	0,93	72,95	233,74	0,66	0,92	72,95	233,74	0,66	0,92	
P18	82,45	189,88	0,55	0,91	71,13	214,13	0,68	0,90	70,41	187,69	0,69	0,90	66,78	218,42	0,69	0,93	73,01	234,28	0,68	0,92	73,01	234,28	0,68	0,92	
P19	82,61	185,84	0,59	0,91	71,43	221,76	0,36	0,91	70,84	195,35	0,47	0,91	67,31	224,01	0,49	0,93	73,35	237,06	0,52	0,92	73,35	237,06	0,52	0,92	
P20	82,67	196,47	0,20	0,92	71,06	214,23	0,57	0,90	70,37	192,59	0,52	0,90	66,98	221,94	0,51	0,93	73,08	238,72	0,47	0,92	73,08	238,72	0,47	0,92	
P21	82,45	181,29	0,67	0,90	71,20	211,23	0,62	0,90	70,38	191,51	0,58	0,90	66,83	213,68	0,72	0,93	73,03	231,14	0,66	0,92	73,03	231,14	0,66	0,92	

4. Discusión y conclusión

Para la validación factorial del cuestionario se utilizó Análisis Factorial Confirmatorio (AFC), partiendo de un modelo *a priori*, fundamentado en bases teóricas y apoyado en la literatura de varios estudios (Brookhart, 2007; Hattie & Timperley, 2007; Jorba & Sanmartí, 2008; Black & Wiliam, 1998; Wiliam, 2009).

El AFC es una técnica que persigue explicar las covarianzas o correlaciones entre un conjunto de variables observadas, lo que permite contrastar un modelo construido con anticipación por el investigador. A diferencia del análisis factorial exploratoria, el confirmatorio supone que el investigador es capaz de aventurar *a priori* la estructura de los datos en función de un marco teórico por lo que sólo necesita confirmar que esa estructura puede también obtenerse empíricamente (Herrero, 2010). Para este tipo de AFC se utilizan Modelos de Ecuaciones Estructurales (SEM). Los modelos SEM son parte de las técnicas multivariantes y es producto de una evolución de la modelización multiecuacional desarrollada principalmente en la econometría y fusionada con los principios de medición de la psicología y la sociología, el SEM es una técnica multivariante que combina aspectos de la regresión múltiple y análisis del factor para estimar una serie de relaciones de dependencia interrelacionadas simultáneamente (Hair, Anderson, Tatham & Black, 2007)

En esencia se ha confirmado las seis dimensiones propuestas en el cuestionario desarrollado por medio de 21 preguntas. Este nuevo instrumento evidenció validez de constructo al igual que otros instrumentos propuestos en las que buscaban validar cuestionarios y/o escalas que evalúan el desempeño educativo del profesorado (Cervelló, Del Villar, Jiménez, Ramos & Blázquez, 2003; Mestanza, 2010; Backhoff & González-Montesinos, 2011).

Por otro lado, en relación a la fiabilidad, por lo general se utiliza el alpha de cronbach para analizar la confiabilidad, en ese sentido, se define como una medida de consistencia interna y su ventaja permite no dividir en dos mitades a los ítems del instrumento, simplemente se aplica la medición y se calcula el coeficiente. Consideramos como criterio aceptable el valor mínimo de 0.70 para la comparación entre grupos y 0.90 para escalas, según lo sugerido por Hernández, Fernández y Baptista (2014, p. 295). De hecho, el presente instrumento evenció valores superiores a lo recomendado, por lo que garantiza una adecuada capacidad de reproductibilidad del instrumento, al menos en este tipo de muestra.

Este instrumento propuesto cuenta con seis dimensiones. La primera dimensión se define como la evaluación formativa asociada a la calificación y está relacionada con las acciones de retroalimentación que realiza el docente después de calificar, utilizando los resultados o el instrumento formativamente en el contexto del aula (Wiliam & Black, 1996). Retroalimentar las pruebas permite abrir la comunicación con los padres entregándoles referentes claros en torno a los resultados de sus hijos, permitiéndoles interpretar un puntaje o calificación. (Moreno, 2016)

La segunda dimensión, la evaluación formativa proactiva, está relacionada con la evaluación anticipada, previa a un proceso de aprendizaje, con el fin de evitar el error o resolver los obstáculos antes de iniciar la tarea. (Jorba & Sanmartí, 2008; Wiliam, 2009). Se promueve que los estudiantes estén provistos de una visión clara de los objetivos de su trabajo y lo que significa lograrlos. Es una instancia para debatir en torno a las instrucciones y criterios a través de la clarificación, comunicación comprensiva de las metas de aprendizaje (Wiliam & Thompson, 2017).

La tercera dimensión, la evaluación formativa interactiva, se realiza durante el proceso de acompañamiento en la producción de conocimiento por parte del estudiante, en donde se monitorea, se recoge información y se retroalimenta en el momento para apoyar el proceso. Según Wiliam (2009) es necesario gestionar actividades que permitan identificar las dificultades de los estudiantes y orientar al sujeto para que reconstruya la respuesta. Esto último, lleva al docente a evaluar su propia práctica, hacer que el estudiante tome conciencia del error o logro y potenciar la gestión de errores para el trabajo cognitivo (Jorba & Sanmartí, 2008)

La cuarta dimensión, la evaluación formativa metacognitiva, está asociada a los procesos internos de la información y su autorregulación, cuyo propósito es propiciar de criterios e instancias reflexivas para que los estudiantes autorregulen su propio proceso de aprender (Hattie & Timperley, 2007). Lo anterior, implica evaluar los procedimientos internos para brindar información al estudiante en torno a las estrategias y la selección de saberes para construir su propio conocimiento. (Bennett, 2011). La idea es que los estudiantes tomen acciones correctivas sobre su desempeño, las investigaciones demuestran que los estudiantes son más autónomos y mejoran su desempeño en la medida que autorregulan su proceso metacognitivo (Black & Wiliam, 1998; Black, Harrison, Lee, Marshall & Wiliam, 2003; Cauley & Mc Millan, 2010; Osorio & López, 2014).

La quinta dimensión, la evaluación formativa retroactiva, está asociada con aquella retroalimentación que opera directamente sobre el resultado o el producto. Para algunos autores (Torrance & Pryor, 1998; Jodaie, Branch, Farrokh & Zoghi, 2011) este tipo de evaluación permite establecer una acción pedagógica “convergente” que ligada al error permite orientar a los estudiantes respecto al criterio de logro y hacer que tomen conciencia del error, complementado con la ayuda del docente, el estudiante pueda buscar la forma de resolver el problema o llegar a la solución. Esto involucra mucho más que corregir, está interesada en guiar al estudiante para que este identifique sus errores y pueda corregirlos por sí mismo (Evans, 2013).

Finalmente, la evaluación formativa ajustada, sexta dimensión, está vinculada con la inclusión. Este tipo de desempeño evaluativo intenta dar respuesta a las necesidades específicas de los estudiantes a partir de las modificaciones que puede realizar el docente en relación al nivel de desarrollo o la trayectoria escolar del estudiante (Wiliam, 2009). Looney (2005) considera que la evaluación formativa permite identificar las necesidades de los estudiantes y ajustar

apropiadamente la enseñanza. Dentro de esta línea, Broadfoot, Daugherty, Gardner, Gipps, Harlen, James y Stobart (1999) sostienen que el aprendizaje mejora cuando el docente es capaz de ajustar la enseñanza teniendo en cuenta los resultados de la evaluación.

El instrumento elaborado puede ser aplicado en el ámbito escolar con el propósito de evaluar el desempeño del profesorado y observar los nudos críticos que requieren de apoyo. Este tipo de evaluación de desempeño permite que la información pueda ser usada principalmente para facilitar la toma de decisiones, para la asignación de incentivos y otras decisiones de la carrera profesional (Goe, Bell y Little, 2008). Sin embargo, una de las limitaciones de la evaluación de desempeño docente y que pone en riesgo el proceso evaluativo, cuyo propósito es la reflexión y aprendizaje del profesorado, es cuando los resultados de los desempeños del profesorado pueden conllevar consecuencias negativas o punitivas hacia el profesorado. (OECD, 2013)

Finalmente, el estudio refleja algunas desventajas como, por ejemplo, el rango de edad investigado es muy limitado y únicamente se investigó en escolares de colegios públicos. Por lo que es necesario abarcar colegios particulares y semi-particulares, puesto que esta información es relevante. Sin embargo, y a la vez, fomenta el avance científico y promueve la acción investigativa de seguir aplicando el instrumento para construir nuevos modelos basados en ecuaciones estructurales y poder comparar la factibilidad de dichos modelos con otras propuestas. Futuros estudios podrían optar por validar con otras técnicas, como la validez de criterio, con lo cual, es posible verificar su validación externa. También, el estudio muestra algunas fortalezas, dado que es uno de los primeros estudios que se interesa por la evaluación formativa durante las prácticas profesionales y este instrumento propuesto podría ser usado para analizar la evaluación formativa de los docentes.

En conclusión, el instrumento desarrollado evidenció propiedades psicométricas adecuadas en cuanto a la validez para evaluar el desempeño docente asociado a las prácticas profesionales. Además, la fiabilidad mostró valores elevados de reproductibilidad en sus seis dimensiones. Los resultados indican que es posible su uso y aplicación a la escuela, las instituciones de formación de profesores y la política pública, con el propósito de abrir la discusión en torno a los procesos de formación y que tipos de práctica deben prevalecer en la formación docente y a la vez permita a las escuelas reflexionar y tomar decisiones en torno al quehacer pedagógico evaluativo formativo para mejorar las prácticas del profesorado que se encuentra en ejercicio.

5. Referencias bibliográficas

- Bennett, R. E. (2011). Formative Assessment: A critical review. *Assessment in Education: Principles, Policy and Practice*, 18(1), 5-25.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, 5(1), 7-74. <https://doi.org/10.1080/0969595980050102>
- Black, P., Harrison, C., & Lee, C. (2003). *Assessment for learning: Putting it into practice*. McGraw-Hill Education (UK).
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment educational assessment, *Evaluation, and accountability* 21(5).
- Backhoff, E., & González-Montesinos, M. J. (2012). Evidencias de validez del cuestionario para docentes del Estudio Internacional Sobre Enseñanza y Aprendizaje (TALIS 2009). *Bordón*, 64(2), 173-194.
- Broadfoot, P. M., Daugherty, R., Gardner, J., Gipps, C. V., Harlen, W., James, M., et al., (1999). *Assessment for learning: Beyond the black box*. Cambridge, UK: University of Cambridge School of Education.
- Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational measurement: Issues and practice*, 22(4), 5-12.
- Brookhart, S. M. (2007). Editorial. *Educational Measurement: Issues and Practice*, 26(3), 1-2. Doi:10.1111/j.1745-3992.2007.00095.x
- Brookhart, S. M. (2009). "Editorial", *Educational Measurement: Issues and Practice*, vol. 28(1), 1-2.
- Carlos-Guzmán, J. (2016). ¿Qué y cómo evaluar el desempeño docente? Una propuesta basada en los factores que favorecen el aprendizaje. *Propósitos y Representaciones*, 4(2), 285-358. Doi: <http://dx.doi.org/10.20511/pyr2016.v4n2.124>
- Cauley, K. y McMillan, J. (2010). Formative Assessment Techniques to Support Student Motivation and Achievement. *The Clearing House: A Journal of Educational Strategies*, 83(1), 1-6.
- Cervelló, E., Del Villar, F., Jiménez, R., Ramos, L., & Blázquez, F. (2003). Clima motivacional en el aula, criterios de éxito de los discentes y percepción de igualdad de trato en función del género en las clases de Educación Física. *Enseñanza*, 21, 379-395.
- EQUIPO, D. T. (2015). *Hacia un sistema completo y equilibrado de evaluación de los aprendizajes en Chile*. Informe Equipo de Tarea para la revisión del SIMCE. Santiago de Chile.
- Evans, C. (2013). Making Sense of Assessment Feedback in Higher Education. *Review of Educational Research*, 83 (1), 70-120. Doi:10.3102/0034654312474350. Disponible en: <http://eprints.soton.ac.uk/373281/1/70.full.pdf>
- García, J., & Medécigo, A. (2014). Los criterios que emplean los estudiantes universitarios para evaluar la ineficacia docente de sus profesores. *Perfiles Educativos*, 26(143). Recuperado de <http://www.revistas.unam.mx/index.php/perfiles/article/view/44026/39835>
- Goe, L., Bell, C. y Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. Washington: National Comprehensive Center for Teacher Quality.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2007). *Multivariate Data Analysis*. 6 Pearson Education Inc.

- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Herrada, M. (2015). Sistema de evaluación del desempeño docente en Chile ¿Mejoran sus resultados los docentes que han sido evaluados en el 2009 y nuevamente el 2013?. Colmee Mexico 2015.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2014). Selección de la muestra. *Metodología de la Investigación (6ª ed., pp. 170-191)*. México: McGraw-Hill.
- Herrero, J. (2010). El Análisis Factorial Confirmatorio en el estudio de la Estructura y Estabilidad de los Instrumentos de Evaluación: Un ejemplo con el Cuestionario de Autoestima CA-14. *Psychosocial Intervention*, 19(3), 289-300.
- Jorba, J., & Sanmartí, N. (2008). *La función pedagógica de la evaluación*. En M. Ballester, Evaluación como un ayuda para el aprendizaje (págs. 21-42). Barcelona: Grao.
- Jodaie, M., Branch, T., Farrokhi, F., & Zoghi, M. (2011). A Comparative Study of EFL [English as a Foreign Language] Teachers' and Intermediate High School Students' Perceptions of Written Corrective Feedback on Grammatical Errors. *English Language Teaching*, 4(4), 36-48. Doi:10.5539/elt.v4n4p36
- Looney, J. (Ed.). (2005) *Formative assessment: improving learning in secondary classrooms*. Paris, France: Organization for Economic Cooperation and Development.
- Lukas, J. y Santiago, K. (2009). *Evaluación educativa*. Madrid: Alianza Editorial.
- Manzi, J., Gonzalez, R. & Sun, Y. (2011). *La Evaluación Docente en Chile*. Centro de Medición MIDE UC. Santiago, Chile. Disponible en: <http://www.mideuc.cl/libroed/>
- Mestanza, J. (2010). Propuesta de evaluación de la actividad docente universitaria en entornos virtuales de aprendizaje. *Revista Española De Pedagogía*, 68(246), 261-279. Retrieved from <http://www.jstor.org/stable/23766300>
- Moreno, T. (2016). *Evaluación del aprendizaje y para el aprendizaje: reinventar la evaluación en el aula*. Universidad Autónoma Metropolitana.
- Núñez, F., & Quiñones, A. (2012). *Importancia de la evaluación y autoevaluación en el rendimiento académico*. Zona Próxima, (16), 96-104. Disponible en: <http://www.redalyc.org/articulo.oa?id=85323935009>
- OECD (2013). *Teachers for the 21st Century*. Using evaluation to improve teaching. París: OECD.
- Osorio, K. y López, A. (2014). La retroalimentación formativa en el proceso de enseñanza-aprendizaje de estudiantes en edad preescolar. *Revista Iberoamericana de Evaluación Educativa*, 7 (1), 13-30.
- Perrenoud, P. (1998). La transposition didactique à partir de pratiques: des savoirs aux compétences. *Revue des sciences de l'éducation*, 24(3), 487-514.
- Perrenoud, P. (2004). *Diez nuevas competencias para enseñar*. Barcelona: Graó.
- Ravela, P., Leymonié, J., Viñas, J., & Haretche, C. (2014). La evaluación en las aulas de secundaria básica en cuatro países de América Latina. *Propuesta educativa*, (41), 20-45.
- Rueda, M., Luna, E., García, B., & Loredó, J. (2010). La evaluación de la docencia en las universidades públicas mexicanas: Un diagnóstico para su comprensión y mejora. *Revista Iberoamericana de Evaluación Educativa*, 3 (1). Recuperado de <https://repositorio.uam.es/handle/10486/661585>
- Shepard, L. A., (2007) *Formative assessment: Caveat emptor*. In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning*. Mahwah, New York: Lawrence Erlbaum Associates.
- Shepard, L., Hammerness, K., Darling-Hammond, L., y Rust, F. (2005). Assessment. En L. Darling-Hammond y J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do*, pp. 275-326. San Francisco: Jossey-Bass.
- Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom*. Inglaterra: McGraw-Hill Education.
- Unicef. (2016). *La naturaleza para el aprendizaje: Usando la investigación para inspirar la práctica*. Panamá: Tinto estudio S.A.
- William, D. (2009) Una síntesis integradora de la investigación e implicancias para una nueva teoría de la evaluación formativa. *Archivos de Ciencias de la Educación* (4a. época), 3(3). Recuperado de: http://www.fuentesmemoria.fahce.unlp.edu.ar/art_revistas/pr.4080/pr.4080.pdf
- William, D. & Black, P. (1996). Meanings and consequences: a basis for distinguishing formative and summative functions of assessment?. *British educational research journal*, 22(5), 537-548.
- William, D., & Thompson, M. (2017). Integrating assessment with learning: What will it take to make it work?. In *The future of assessment* (pp. 53-82). Routledge.