

Artículo de Investigación



Escala para evaluar la Literacidad Evaluativa en Docentes de Educación Básica

Scale to assess Evaluative Literacy in Basic Education Teachers

José Pérez García*

Centro Universitario CIFE

resiljpg@gmail.com

<https://orcid.org/0000-0002-7824-7535>

María de los Ángeles López Ortega

Centro Universitario CIFE

alopez25@cife.edu.mx

<https://orcid.org/0000-0002-1435-8009>

Sección: **Artículo de investigación**

Fecha de recepción: **8/10/2022** | Fecha de aceptación: **23/12/2022**

Referencia del artículo en estilo APA 7^a. edición:

Pérez García, J., & López Ortega, M. A. (2020). Escala para evaluar la Literacidad Evaluativa en Docentes de Educación Básica. *Transdigital*, 3(6). 1–27.
<https://doi.org/10.56162/transdigital159>



Licencia [Creative Commons Attribution 4.0](https://creativecommons.org/licenses/by/4.0/)

International License (CC BY 4.0)

Resumen

Los docentes requieren una formación consistente en el campo de evaluación para obtener información válida y confiable sobre el aprendizaje de los alumnos. Con esta finalidad se diseñó y validó una escala para medir la Literacidad Evaluativa de Docentes de Educación Básica. En este trabajo se propone una estructura del constructo centrada en cuatro dimensiones: Conocimiento y uso de evaluaciones estandarizadas, Asignación y uso de la calificación, Uso de la evaluación, y Evaluación formativa. Se aplicó el instrumento a una muestra de 302 docentes. Se realizó Análisis Factorial Exploratorio y Análisis Factorial Confirmatorio para confirmar el modelo teórico en el que se basa, y se obtuvieron los índices de confiabilidad y ajuste apropiados para demostrar la validez del instrumento: Alfa de Cronbach 0,96 y confiabilidad compuesta 0,95. Se puede concluir que el instrumento es fiable y válido para diagnosticar el nivel de literacidad evaluativa de docentes de educación básica.

Palabras clave: evaluación del aprendizaje, literacidad, profesores, educación básica.

Abstract

Teachers require consistent training in the field of assessment to obtain valid and reliable information about student learning. For this purpose, a scale was designed and validated to measure the Evaluative Literacy of Basic Education Teachers. In this work, a structure of the construct centered on four dimensions is proposed: Knowledge and use of standardized evaluations, Assignment and use of the qualification, Use of the evaluation, and Formative evaluation. The instrument was applied to a sample of 302 teachers. Exploratory Factor Analysis and Confirmatory Factor Analysis were performed to confirm the theoretical model on which it is based, and the appropriate reliability and fit indices were obtained to demonstrate the validity of the instrument: Cronbach's Alpha 0.96 and composite reliability 0.95. It can be concluded that the instrument is reliable and valid for diagnosing the evaluative literacy level of basic education teachers.

keywords: learning assessment, literacy, teachers, basic education.

1. Introducción

Uno de los aspectos críticos de la actividad docente es la evaluación del desempeño de sus alumnos. Y no siempre se sienten cómodos asumiendo esta función, pues piensan que no han recibido la formación que les permite tomar las mejores decisiones al respecto (Mertler, 2009). Para realizar esta actividad se requiere un conjunto de habilidades y conocimientos, aunque esto no es suficiente, ya que esta se realiza en un contexto social influido por un currículo, orientaciones pedagógicas y las expectativas de la comunidad (Looney y colaboradores, 2018).

Un educador profesional con conocimientos sobre evaluación debería saber cómo obtener información confiable sobre el aprendizaje de sus alumnos mediante evaluaciones de alta calidad. También debería saber cómo promover y certificar el aprendizaje a partir de los resultados del proceso de evaluación (Stiggins, 2017). Esta capacidad se conoce como literacidad en evaluación, y refiere al conjunto de conocimientos y habilidades que tiene el docente en el campo de la evaluación, que le permiten medir y apoyar el aprendizaje de los alumnos (Brookhart, 2011) y tiene como constructo próximo o sustituto el de competencia evaluativa (DeLuca et al., 2016). También incluye la capacidad que tiene el docente para adquirir información en diferentes medios relacionada con la práctica de la evaluación, analizar y valorar su pertinencia; utilizarla para la mejora de su práctica y saber comunicar a diferentes audiencias los resultados de sus evaluaciones (OECD, 2019). Por lo tanto, la persona con literacidad evaluativa sabe qué método de evaluación utilizar para el acopio de información pertinente sobre el desempeño de los alumnos, sabe cómo maximizar el aprendizaje a partir de esa información y sabe comunicar los resultados (Huang y He, 2016).

En 1990 hubo una de las contribuciones más significativas al tema de la literacidad evaluativa con la publicación de los Estándares para la Competencia Docente en Evaluación Educativa de Estudiantes. Estos establecen que el docente debe tener competencias para diseñar y seleccionar métodos de evaluación adecuados; administrar, calificar e interpretar los resultados, tomar decisiones a partir de los resultados; desarrollar procedimientos válidos de calificación, compartir los resultados e identificar casos no éticos de evaluación (Coombe et al., 2020). Estos estándares fueron un primer referente importante. Sin embargo,

como establece Brookhart (2011), la evaluación formativa, la evaluación basada en estándares y la rendición de cuentas hace necesario actualizar dichos estándares.

Uno de los primeros dispositivos para evaluar la literacidad evaluativa fue elaborado por Plake y colaboradores en 1993 (Teacher Assessment Literacy Questionnaire). En la primera parte contiene 35 reactivos que miden las siete áreas de competencia consideradas en los estándares de 1990 (Standards for Teacher Competence in Educational Assessment of Students [AFT, NCME, NEA], 1990). Para su elaboración se utilizó una tabla de especificaciones donde se estableció el nivel de conocimientos requerido en cada una de las áreas de competencia. En las áreas de competencia se refiere que los profesores deben estar capacitados para: 1) Elegir métodos de evaluación apropiados; 2) Desarrollar métodos de evaluación apropiados; 3) Analizar, calificar e interpretar los resultados tanto de evaluaciones externas como de las que elabora él mismo; 4) Utilizar los resultados de las evaluaciones en la toma de decisiones para la mejora; 5) Desarrollar procedimientos válidos para calificar a los alumnos; 6) Comunicar los resultados de la evaluación a diversas audiencias y 7) Reconocer métodos de evaluación no éticos, inapropiados e ilegales. La segunda parte del instrumento contiene preguntas relacionadas con los antecedentes y percepciones de los docentes sobre las pruebas.

En Etiopía se utilizó una versión modificada del instrumento elaborado por Plake y colaboradores (1993). En este caso, el instrumento se dividió en dos partes. La primera para el acopio de información del contexto y la segunda está formada por 31 reactivos de opción múltiple con cuatro opciones de respuesta. Se eliminaron cuatro reactivos del instrumento original debido a que no son relevantes en el sistema educativo de Etiopía (Tagele y Bedilu, 2015).

Otro instrumento elaborado referido a la práctica de evaluación de los profesores es el de Zhang y Burry-Stock (1997). Este instrumento, denominado *Assessment Practices Inventory* (API), contiene 67 preguntas con una escala de cinco puntos en donde el profesor contesta cómo percibe su competencia en evaluación. A partir del Análisis Factorial se identificaron las dimensiones subyacentes: 1) Desarrollar y administrar pruebas de lápiz y papel o elegir pruebas; 2) Interpretar los resultados de las pruebas estandarizadas, calcular

las estadísticas de las pruebas y usar los resultados en la toma de decisiones; 3) Desarrollar y utilizar evaluaciones de desempeño, evaluaciones informales; 4) Comunicar los resultados de las pruebas; 5) Calificación basada en falta de logros; 6) Ética de la evaluación; y, 7) Calificación.

La encuesta desarrollada para docentes de primaria y secundaria *Measurement Literacy Questionnaire*, fue elaborada por Daniel y King (1998). Está formada por 67 reactivos, de los cuales siete se refieren a información de contexto, 30 aluden a la literacidad referida a pruebas y medición, y los 30 restantes refieren al uso de técnicas de medición.

Otro instrumento elaborado a partir de los estándares de 1990 es el *Assessment Literacy Inventory* (ALI). Mertler y Campbell (2005) redactaron 35 reactivos integrados en 5 escenarios que se basan en las decisiones que el profesor enfrenta en el aula con respecto a la evaluación. De cada escenario se derivan siete reactivos de opción múltiple alineados con los estándares de 1990, ya mencionados en el estudio de Plake y colaboradores (1993). Este instrumento se utilizó por Hailaya y colaboradores (2014), como una versión adaptada al contexto filipino, hicieron adaptaciones en la descripción de los escenarios, temas y nombres de los maestros y, para mantener la integridad del instrumento, conservaron la estructura original de los escenarios.

Mediante una metodología mixta, Yamtim y Wongwanich (2014) abordaron este tema utilizando el método del grupo de enfoque y el cuestionario *Classroom Assessment Literacy Inventory* (CALI), que es una adaptación del que utilizó Plake en 1993. Incluyeron en la primera parte del cuestionario datos sociodemográficos y en la segunda parte hay 35 reactivos, con cuatro opciones de respuesta, referidos a la literacidad evaluativa en el aula. Se utilizó un protocolo para el grupo de enfoque.

Alkharusi y colaboradores (2014) elaboraron un cuestionario de autoreporte para maestros de los grados 5º. al 12º. del sultanato de Omán. La primera parte recoge información demográfica y de contexto. La segunda parte contiene información referida a seis dimensiones: 1) Actitudes hacia la evaluación educativa; 2) Competencia autopercebida en evaluación educativa que tiene a su vez cuatro componentes: desarrollo y administración

de métodos de evaluación, análisis de resultados de evaluación, desarrollo y calificación de evaluación del desempeño, desarrollo de procedimientos de calificación y comunicación de resultados a diversas audiencias. Adicionalmente, los autores agregaron siete reactivos referidos a los asuntos éticos de la evaluación. Las respuestas se dieron en una escala Likert de 5 puntos; 3) Conocimiento en evaluación educativa; 4) Prácticas de evaluación educativa, que se divide en los siguientes segmentos: métodos tradicionales de evaluación, métodos alternativos de evaluación, análisis de resultados de evaluación, comunicación de la evaluación, estándares y criterios de evaluación, involucramiento del estudiante en la evaluación y factores de calificación no asociados al logro; 5) Uso de pruebas en el aula, la cual exploró 10 áreas de decisiones académicas; 6) Actitud hacia las pruebas en el aula.

Dos estudios más recientes sobre la elaboración de instrumentos para evaluar la literacidad evaluativa (Iqbal et al., 2020) consisten en una prueba de opción múltiple para evaluar conocimientos relacionados con cuatro dimensiones: tipos de evaluación, uso de los resultados de la evaluación, teorías de medición y lo que establecen los estándares profesionales nacionales para maestros del gobierno de Pakistán. El otro estudio es el de Nikmard y Mohamadi Zenouzagh (2020), que utiliza una metodología mixta, con cuatro preguntas abiertas y un cuestionario de 25 reactivos con escala tipo Likert de cinco puntos. El análisis factorial reveló la presencia de cuatro componentes: validez, confiabilidad, interpretabilidad de los resultados y eficiencia.

En el caso de México, existe una investigación sobre la importancia y familiaridad de las prácticas de evaluación desde la perspectiva de los maestros, orientadores y administradores educativos. El cuestionario se elaboró considerando las siete áreas de competencias señaladas en los estándares de 1990, competencias de orientación profesional y prácticas de evaluación relevantes para México (Arce-Ferrer, Cab, y Cisneros-Cohernour, 2001). De los 74 reactivos que contiene el instrumento, 53 son de traducción directa de competencias, nueve de la adaptación cultural de las competencias y 12 referidos a prácticas de evaluación específicas de México.

Existen otros dos estudios que, aunque no tienen por objeto explorar únicamente la literacidad evaluativa, consideran algún componente referido a este tema. El primero es de

corte cuantitativo, en el cual se elaboró un cuestionario para profesores de secundaria en México. Se obtuvieron datos generales de los participantes e información sobre competencias docentes en la asignatura de tecnología. Una de las competencias que exploraron se refiere a la evaluación (Morán y colaboradores 2015).

El segundo estudio utilizó una metodología mixta, que consistió en una entrevista semiestructurada a docentes en México y un cuestionario dirigido a alumnos (Ramón et al., 2017). Los investigadores elaboraron una guía para la entrevista con 33 preguntas abiertas que corresponden a siete categorías; una de las cuales se refiere al empleo de la evaluación con enfoque formativo (seis preguntas). El cuestionario para los alumnos contiene 16 preguntas con escala tipo Likert, de las cuales cinco se refieren al empleo de la evaluación con enfoque formativo.

Finalmente, la Secretaría de Educación Pública publicó un documento (SEP, 2019), donde se plantea uno de los dominios que el docente debe tener: “Evalúa de manera permanente el desempeño de los alumnos, a través de diversas estrategias para valorar los aprendizajes y su intervención docente” (SEP, 2019. p. 22). Para ello, señala que el maestro: “Utiliza estrategias de evaluación diversificadas, permanentes y flexibles, coherentes con los aprendizajes que espera lograr, las actividades didácticas realizadas y las características de los alumnos” (SEP, 2019. p. 22).

Las evaluaciones externas que recogen información sobre el rendimiento de los estudiantes empezaron a tener mayor presencia en México al inicio de este siglo, lo cual permitió que los docentes tuvieran la oportunidad de invertir mucho tiempo en el análisis de información derivada de evaluaciones estandarizadas y, para realizar bien esta actividad requerían competencias específicas (Arce-Ferrer, Cab, y Cisneros-Cohernour, 2001).

Los cambios en el campo de la evaluación expresados en documentos que orientan y regulan las prácticas docentes hacen necesario contar con información válida y confiable que nos permita conocer cómo son las prácticas actuales de evaluación que el docente realiza en el aula y, en función de la información obtenida, orientar decisiones específicas en la formación, actualización y capacitación docente. Los propósitos del presente estudio son:

1) Diseñar un instrumento pertinente y práctico para evaluar la literacidad evaluativa de docentes de educación básica en México; 2) Realizar la validación de contenido del instrumento con un grupo de jueces, para determinar su grado de relevancia y coherencia teórica; 3) Ejecutar un análisis factorial para valorar sus dimensiones; y, 4) Analizar la confiabilidad del instrumento mediante la aplicación del Alfa de Cronbach.

2. Método de investigación

2.1. Tipo de estudio

Se realizó un estudio instrumental de validez y confiabilidad respecto al instrumento *Escala para Medir Literacidad Evaluativa en Docentes*. Este tipo de estudio consiste en determinar las propiedades psicométricas de nuevos instrumentos (Montero y León, 2007).

2.2. Participantes

Una muestra de 302 docentes (196 mujeres y 106 hombres) de escuelas públicas de educación básica, con experiencia docente frente a grupo de uno hasta 53 años ($X=16.57$; $SD=10.55$) seleccionados mediante un muestreo no probabilístico, con el criterio de que se encontraran desarrollando actividades frente a grupo (Tabla 1).

Tabla 1

Muestra de profesores participantes

Nivel educativo	N	Porcentaje de la muestra
Preescolar	21	7,0
Primaria	75	24,8
Secundaria	206	68,2
Total	302	100,0

2.3. Instrumento

Se utilizó un instrumento para evaluar la *Literacidad Evaluativa en los Docentes* (LED) y se complementó con el *Cuestionario de Factores Sociodemográficos* elaborado por el CIFE

(2015), con el cual se hizo acopio de información sociodemográfica de los docentes, tales como: edad, sexo, grado de estudio, años de experiencia docente, nivel económico y capacitación recibida.

2.4. Procedimiento

El estudio de validez y confiabilidad del instrumento se llevó a cabo mediante las siguientes fases.

2.5. Diseño y revisión por expertos

Se diseñó el instrumento mediante una escala. El instrumento está compuesto por las siguientes dimensiones, de acuerdo a la revisión de la literatura: 1) Seleccionar y desarrollar métodos de evaluación; 2) Administrar, calificar e interpretar resultados; 3) Utilizar los resultados para las decisiones cotidianas; 4) Comunicar los resultados a otros; 5) Usar éticamente la evaluación; 6) Evaluar de manera formativa; y 7) Prácticas de calificación. Posee 44 ítems (en la Tabla 2 se describen de manera sintética las dimensiones y el número de ítems por dimensión). La forma de calificación o evaluación es mediante una escala con la cual el docente estima sus conocimientos y habilidades referidos a cada tarea descrita en los ítems. El nivel de la escala va de 1: *muy bajo*; hasta, 5: *muy alto*.

Tabla 2

Dimensiones y preguntas del instrumento

Dimensiones	Ítems
1. Selección y desarrollo de métodos de evaluación	3,11,15,17,18,27,29
2. Administrar, calificar e interpretar resultados	2,8,13,14,19,25,26
3. Utilizar los resultados para las decisiones cotidianas	4,6,9,22,23,24
4. Comunicar los resultados a otros	1,7,16,21,30
5. Uso ético de la evaluación	5,10,12,20,28
6. Evaluación Formativa	31,32,33,34,35,36
7. Prácticas de calificación	37,38,39,40,41,42,43,44

2.6. Estudio de la validez de contenido

Una vez que el instrumento fue elaborado con el apoyo de expertos, se realizó un estudio sobre la validez de contenido mediante la evaluación por parte de 11 jueces mediante la V de Aiken a través de una plantilla de Excel. Esta técnica consiste en solicitar la opinión de personas cuya trayectoria en el tema hace que se les reconozca como expertas y, por lo tanto pueden establecer juicios de valor y proporcionar información (Escobar-Pérez y Cuervo-Martínez, 2008). Sus recomendaciones ayudan en la elaboración del instrumento de medición.

Se envió invitación a 24 expertos para participar en la validación de contenido, considerando como criterios para su selección su experiencia en la evaluación de instrumentos, experiencia en investigación y realización de publicaciones, así como la disponibilidad para participar (Tabla 3). Los 11 jueces que aceptaron evaluaron la pertinencia y redacción de cada uno de los ítems del instrumento.

Tabla 3

Datos de la competencia de los jueces

		Frecuencia	Porcentaje
Género	Hombres	6	54.5%
	Mujeres	5	45.5%
Educación	Doctorado completo	10	90.9%
	Doctorado incompleto	0	0.0%
	Maestría completa	1	9.1%
	Especialización	0	0.0%
		Media	Desviación estándar
Edad		46.09	7.53
Conferencias por invitación		6.64	3.26
Grado de experiencia como revisor (1-4)		3.45	1.16
Años de experiencia	Docencia	14.73	7.71
	Investigación	9.00	7.22
Publicaciones	Artículos	10.27	11.06
	Libros	0.27	0.62
	Capítulos	3.36	4.92
	Ponencias	7.64	7.15
Educación continuada	Cursos	23.64	17.63

Para la evaluación cuantitativa se utilizó una escala Likert con valores de uno a cuatro para que el experto valorara la relevancia y redacción del ítem. Para el análisis cuantitativo se obtuvo la media, considerando aceptable un valor mayor a tres; el índice de validez se calculó utilizando la V de Aiken y se aceptaron valores superiores a 0.75 con un $p < 0,05$. Los ítems con valores inferiores fueron eliminados de la prueba (Tabla 4).

Tabla 4

Resultados de la validez de contenido por los jueces (n = 11)

Ítem	Variable	Media	D.E.	V de Aiken*	Límite Inferior*	Límite Superior
Ítem 1	Pertinencia	3.36	0.98	0.788	0.622	0.893
	Redacción	3.27	0.62	0.758	0.590	0.872
Ítem 2	Pertinencia	3.09	1.00	0.697	0.527	0.826
	Redacción	3.09	0.90	0.697	0.527	0.826
Ítem 3	Pertinencia	3.64	0.64	0.879	0.727	0.952
	Redacción	3.45	0.66	0.818	0.656	0.914
Ítem 4	Pertinencia	3.64	0.64	0.879	0.727	0.952
	Redacción	3.64	0.48	0.879	0.727	0.952
Ítem 5	Pertinencia	3.36	0.64	0.788	0.622	0.893
	Redacción	3.55	0.50	0.848	0.691	0.933
Ítem 6	Pertinencia	3.55	0.50	0.848	0.691	0.933
	Redacción	3.36	0.48	0.788	0.622	0.893
Ítem 7	Pertinencia	3.55	0.50	0.848	0.691	0.933
	Redacción	3.18	0.57	0.727	0.558	0.849
Ítem 8	Pertinencia	3.82	0.39	0.939	0.804	0.983
	Redacción	3.55	0.50	0.848	0.691	0.933
Ítem 9	Pertinencia	3.36	0.64	0.788	0.622	0.893
	Redacción	3.18	0.72	0.727	0.558	0.849
Ítem 10	Pertinencia	3.64	0.48	0.879	0.727	0.952
	Redacción	3.64	0.48	0.879	0.727	0.952
Ítem 11	Pertinencia	3.64	0.88	0.879	0.727	0.952

Tabla 4

Resultados de la validez de contenido por los jueces (n = 11)

Ítem	Variable	Media	D.E.	V de Aiken*	Límite Inferior*	Límite Superior
Ítem 12	Redacción	3.18	1.11	0.727	0.558	0.849
	Pertinencia	3.64	0.64	0.879	0.727	0.952
Ítem 13	Redacción	3.55	0.50	0.848	0.691	0.933
	Pertinencia	3.18	1.03	0.727	0.558	0.849
Ítem 14	Redacción	3.55	0.50	0.848	0.691	0.933
	Pertinencia	3.91	0.29	0.970	0.847	0.995
Ítem 15	Redacción	3.64	0.48	0.879	0.727	0.952
	Pertinencia	3.73	0.45	0.909	0.764	0.969
Ítem 16	Redacción	3.55	0.50	0.848	0.691	0.933
	Pertinencia	3.36	1.15	0.788	0.622	0.893
Ítem 17	Redacción	3.27	0.86	0.758	0.590	0.872
	Pertinencia	3.91	0.29	0.970	0.847	0.995
Ítem 18	Redacción	3.45	0.89	0.818	0.656	0.914
	Pertinencia	3.55	0.66	0.848	0.691	0.933
Ítem 19	Redacción	3.64	0.48	0.879	0.727	0.952
	Pertinencia	3.45	0.99	0.818	0.656	0.914
Ítem 20	Redacción	3.45	0.89	0.818	0.656	0.914
	Pertinencia	3.45	0.99	0.818	0.656	0.914
Ítem 21	Redacción	3.73	0.45	0.909	0.764	0.969
	Pertinencia	3.91	0.29	0.970	0.847	0.995
Ítem 22	Redacción	3.91	0.29	0.970	0.847	0.995
	Pertinencia	3.64	0.88	0.879	0.727	0.952
Ítem 23	Redacción	3.73	0.45	0.909	0.764	0.969
	Pertinencia	3.91	0.29	0.970	0.847	0.995
Ítem 24	Redacción	3.82	0.39	0.939	0.804	0.983
	Pertinencia	4.00	0.00	1.000	0.896	1.000
Ítem 25	Redacción	3.64	0.48	0.879	0.727	0.952
Ítem 25	Pertinencia	3.91	0.29	0.970	0.847	0.995

Tabla 4

Resultados de la validez de contenido por los jueces (n = 11)

Ítem	Variable	Media	D.E.	V de Aiken*	Límite Inferior*	Límite Superior
Ítem 26	Redacción	3.82	0.39	0.939	0.804	0.983
	Pertinencia	4.00	0.00	1.000	0.896	1.000
Ítem 27	Redacción	3.82	0.39	0.939	0.804	0.983
	Pertinencia	4.00	0.00	1.000	0.896	1.000
Ítem 28	Redacción	3.82	0.39	0.939	0.804	0.983
	Pertinencia	3.55	0.66	0.848	0.691	0.933
Ítem 29	Redacción	3.91	0.29	0.970	0.847	0.995
	Pertinencia	3.64	0.64	0.879	0.727	0.952
Ítem 30	Redacción	4.00	0.00	1.000	0.896	1.000
	Pertinencia	3.73	0.45	0.909	0.764	0.969
Ítem 31	Redacción	3.64	0.88	0.879	0.727	0.952
	Pertinencia	3.82	0.39	0.939	0.804	0.983
Ítem 32	Redacción	3.73	0.62	0.909	0.764	0.969
	Pertinencia	3.64	0.48	0.879	0.727	0.952
Ítem 33	Redacción	3.64	0.88	0.879	0.727	0.952
	Pertinencia	3.73	0.45	0.909	0.764	0.969
Ítem 34	Redacción	3.91	0.29	0.970	0.847	0.995
	Pertinencia	3.73	0.45	0.909	0.764	0.969
Ítem 35	Redacción	3.55	0.66	0.848	0.691	0.933
	Pertinencia	3.73	0.45	0.909	0.764	0.969
Ítem 36	Redacción	3.73	0.86	0.909	0.764	0.969
	Pertinencia	3.64	0.64	0.879	0.727	0.952
Ítem 37	Redacción	3.64	0.64	0.879	0.727	0.952
	Pertinencia	3.64	0.88	0.879	0.727	0.952
Ítem 38	Redacción	3.55	0.78	0.848	0.691	0.933
	Pertinencia	3.64	0.88	0.879	0.727	0.952
Ítem 39	Pertinencia	3.55	0.78	0.848	0.691	0.933

Tabla 4

Resultados de la validez de contenido por los jueces (n = 11)

Ítem	Variable	Media	D.E.	V de Aiken*	Límite Inferior*	Límite Superior
Ítem 40	Redacción	3.45	0.89	0.818	0.656	0.914
	Pertinencia	3.27	1.05	0.758	0.590	0.872
	Redacción	3.55	0.89	0.848	0.691	0.933
Ítem 41	Pertinencia	3.36	0.88	0.788	0.622	0.893
	Redacción	3.73	0.62	0.909	0.764	0.969
Ítem 42	Pertinencia	3.82	0.57	0.939	0.804	0.983
	Redacción	3.82	0.57	0.939	0.804	0.983
Ítem 43	Pertinencia	3.82	0.57	0.939	0.804	0.983
	Redacción	3.64	0.64	0.879	0.727	0.952
Ítem 44	Pertinencia	3.73	0.86	0.909	0.764	0.969
	Redacción	3.82	0.57	0.939	0.804	0.983
Escala Total	Pertinencia	3.45	0.50	0.818	0.656	0.914
	Redacción	3.36	0.48	0.788	0.622	0.893
	Satisfacción	4.00	0.85	0.750	0.606	0.854

Nota: D.E. = Desviación estándar.

2.7. Análisis factorial

Se aplicó el instrumento a una muestra no probabilística de 302 docentes de educación básica. En la Tabla 5 se describen las características de esta muestra.

Tabla 5

Datos sociodemográficos de los participantes

		Frecuencia	Porcentaje
Género	Hombres	106	35.09%
	Mujeres	196	64.90%
Estado Civil	Soltero	84	27.81%
	Casado/unión libre	177	58.60%

Tabla 5

Datos sociodemográficos de los participantes

		Frecuencia	Porcentaje
Condiciones económicas	Divorciado	27	8.94%
	Viudo	14	4.63%
	Muy bajas	3	0.99%
	Bajas	44	14.57%
	Aceptables	186	61.58%
	Buenas	65	21.52%
	Excelentes	4	1.35%
	Preescolar	21	6.95%
Situación laboral	Primaria	75	24.83%
	Secundaria General	36	11.92%
	Secundaria Técnica	170	56.29%
		Media	D.E.
Edad	44.65	10.05	
Años de experiencia como docente	17.88	10.61	

De acuerdo con la revisión teórica, el instrumento se construyó considerando siete factores y, con la finalidad de verificar dicha estructura, se realizó en primer lugar un Análisis Factorial Exploratorio (AFE): el propósito fue analizar los datos obtenidos y encontrar la estructura subyacente expresada en un número pequeño de dimensiones que explicaran la mayor cantidad de varianza posible (Guti, 2019). El análisis factorial se efectuó mediante el programa JASP 14.1. En la Tabla 6 se muestran las cargas factoriales, así como los cuatro factores obtenidos en el AFE. A partir del AFE se eliminaron tres ítems que tuvieron cargas factoriales altas en más de un factor.

El valor estadístico proporcionado por la prueba KMO fue de 0.970, lo que permite concluir que hay una buena adecuación de los datos al modelo de cuatro factores. La prueba de esfericidad de Bartlett fue estadísticamente significativa ($X^2 = 9003.01$, $p = 0.01$).

Tabla 6

Cargas factoriales

Factor	Reactivo	Carga factorial			
		1	2	3	4
1: Asignación y uso de calificaciones					
V36	Asignar calificaciones que reflejen los aprendizajes de los estudiantes.	.681			
V37	Utilizar las calificaciones para mejorar la enseñanza y aprendizaje.	.670			
V38	Utilizar las calificaciones como método para ayudar a los estudiantes.	.710			
V39	Utilizar las calificaciones para motivar a los estudiantes a aprender.	.702			
V40	Tomar en cuenta la opinión de los alumnos para calificar.	.635			
V41	Analizar la distribución de calificaciones de un grupo para tomar decisiones.	.686			
V42	Establecer criterios de calificación pertinentes.	.762			
V43	Asignar una calificación final justa.	.752			
2: Conocimiento y uso de evaluaciones estandarizadas					
V8	Usar los resultados de las evaluaciones acumulativas para ajustar los planes:		.477		
V12	Interpretar las puntuaciones de los informes de resultados de las pruebas estandarizadas:		.581		
V15	Explicar a los padres cómo utilizar los resultados de la evaluación para mejorar:		.517		
V16	Determinar si una evaluación está alineada con los estándares requeridos:		.580		
V17	Evaluaciones externas actualizadas y disponibles:		.765		
V18	Administrar evaluaciones estandarizadas:		.752		
V19	Reconocer cuando se utilizan los resultados de la evaluación de manera inapropiada:		.491		
V22	Usar los resultados del seguimiento de los alumnos para ajustar la planeación:		.490		
V24	Interpretar puntuaciones basadas en criterios:		.526		
V25	Entender e interpretar los resultados de pruebas estandarizadas:		.648		
V27	Identificar mis responsabilidades legales con respecto a la evaluación:		.495		
V28	Seleccionar los aprendizajes pertinentes a evaluar según el propósito:		.484		
3: Evaluación formativa					
V4	Cumplir con los límites de confidencialidad con respecto a los resultados de la evaluación:		.440		
V5	Usar los resultados de la evaluación para formar grupos o equipos:		.473		
V20	Comunicar los resultados de las evaluaciones a los estudiantes:		.586		
V21	Usar la información de la evaluación para desarrollar un plan de enseñanza:		.503		
V23	Usar los resultados de la evaluación para identificar a los estudiantes con necesidades educativas:		.523		

Tabla 6

Cargas factoriales

Factor	Reactivo	Carga factorial			
		1	2	3	4
V26	Desarrollar evaluaciones con diferentes formatos:			.449	
V29	Explicar a los estudiantes cómo se utilizarán los resultados de las evaluaciones para asignar calificaciones:			.481	
V30	Involucrar a los estudiantes en los procesos de evaluación:			.532	
V32	Lograr que mis evaluaciones sean justas, confiables y proporcionen información válida:			.551	
V33	Proporcionar retroalimentación oportuna a los estudiantes:			.715	
V34	Proporcionar comentarios útiles a los estudiantes para mejorar su aprendizaje:			.613	
V35	Involucrar a los estudiantes en monitorear su propio aprendizaje y usar la información de la evaluación:			.599	
4: Uso de la evaluación					
V1	Explicar claramente a los padres los resultados de la evaluación:			.529	
V2	Elegir un método de evaluación para un propósito específico:			.492	
V3	Usar los resultados de la evaluación para ajustar mis estrategias didácticas:			.471	
V7	Explicar los resultados de la evaluación a otros docentes para ayudar a tomar decisiones:			.504	
V9	Consecuencias del uso poco ético de la evaluación:			.435	
V10	Seleccionar múltiples métodos de evaluación (por ejemplo, pruebas, observaciones):			.628	
V11	Reconocer el uso inadecuado de la evaluación:			.498	
V13	Aplicación de evaluaciones de seguimiento del aprendizaje de los alumnos:			.474	

Los cuatro factores pueden considerarse como los cuatro constructos que el instrumento pretende medir. El factor 1: Asignación y uso de la calificación, con 8 ítems; factor 2: Conocimiento y uso de las evaluaciones estandarizadas, que contiene 12 ítems; factor 3: Evaluación formativa, con 12 ítems y factor 4: Uso de la evaluación, con 8 ítems.

Posteriormente, se realizó el Análisis Factorial Confirmatorio (AFC) mediante la técnica de componentes principales. El AFC se efectuó mediante el programa JASP 14.1 para evaluar el ajuste al modelo (Tablas 7 y 8).

Tabla 7

Criterios de ajuste del modelo

Criterios de ajuste	Nivel aceptable	Interpretación
Chi cuadrada	Valor de X2 en la tabla	Compara el valor de χ^2 obtenido con el de la tabla.
Índice de Bondad de Ajuste (GFI)	0 (sin ajuste) a 1 (ajuste perfecto)	Los valores cercanos a .90 o .95 reflejan un buen ajuste
Error cuadrático medio de aproximación (RMSEA)	.05 a .08	El valor de .05 a .08 indica un ajuste perfecto
Índice de ajuste normalizado (NFI)	0 (sin ajuste) a 1 (ajuste perfecto)	El valor cercano a .90 o .95 refleja un buen ajuste

Nota. Adaptado de Khodabakhshzadeh et al. (2018)

Tabla 8

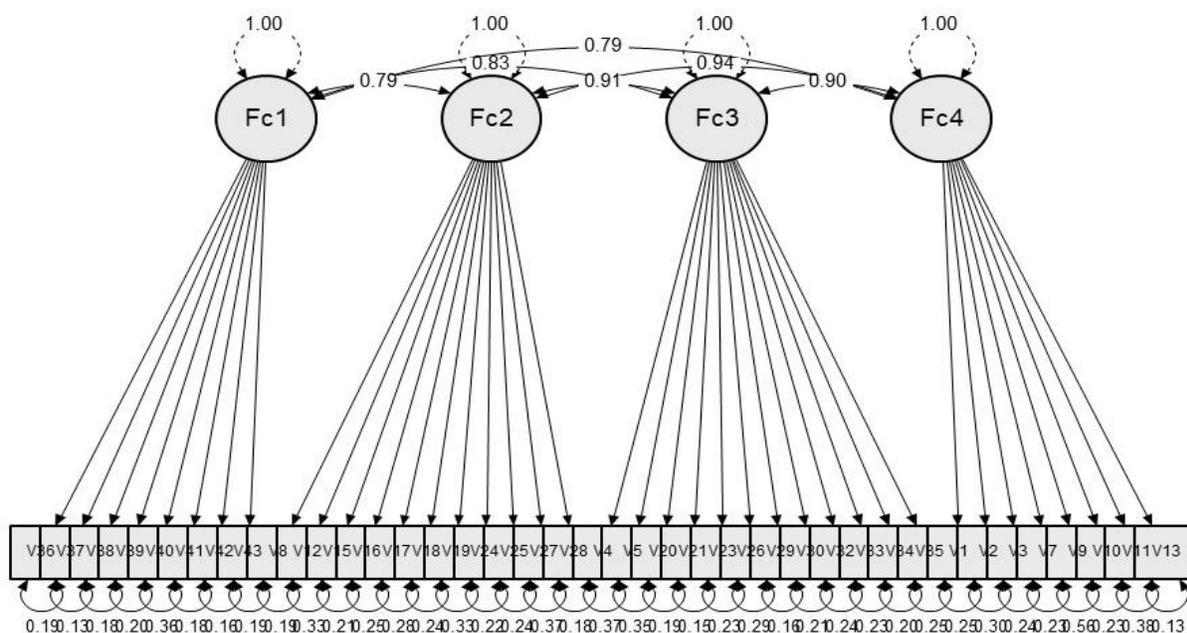
Índices de ajuste del modelo para el instrumento: evaluación de la literacidad evaluativa

	X2	df	X2/df	GFI	NFI	RMSEA
Ajuste aceptable			<3	>.90	>.90	<.08
Modelo	168.820	623	0.270	0.996	0.995	0.000

De acuerdo con la estructura de la tabla el, AFC confirmó la estructura del instrumento. Para examinar la confiabilidad del instrumento, se utilizó el alfa de Cronbach. El coeficiente de confiabilidad fue .981 y una confiabilidad compuesta de 0,95 (Figura 1).

Figura 1

Modelo AFC para escala de evaluación formativa



Nota. Factor 1 = asignación y uso de la calificación; Factor 2 = conocimiento y uso de evaluaciones estandarizadas; Factor 3 = evaluación formativa y factor 4 = uso de la evaluación.

2.8. Aspectos éticos

En la investigación se cumplieron los siguientes criterios éticos de la investigación: 1) Todos los participantes firmaron un consentimiento antes de completar los instrumentos; 2) Los participantes podían retirarse en cualquier momento; y, 3) Hubo protección de los datos personales de todos los participantes acorde con la Ley Mexicana de Protección de Datos Personales.

3. Resultados

3.1. Diseñar un instrumento pertinente y práctico para evaluar la literacidad evaluativa de docentes de educación básica en México

La validación del instrumento se realizó en dos pasos: AFE con rotación varimax y posteriormente el AFC para determinar la robustez de la estructura de factores. Dado que en las escalas tipo Likert se incumplen dos de los supuestos del AFC: la normalidad multivariante de las variables observadas y que estas y las variables latentes sean continuas; se utilizó el método de mínimos cuadrados no ponderados (ULS) (Morata et al., 2015).

Lo anterior permite contar con un instrumento robusto con 40 ítems para valorar los conocimientos y habilidades de los docentes de educación básica referidos a la literacidad evaluativa.

3.2. Realizar la validación de contenido del instrumento con un grupo de jueces, para determinar su grado de relevancia y coherencia teórica

En la Tabla 3 se describieron los datos de validez de contenido mediante la prueba V de Aiken. Pudo observarse que el instrumento posee validez de contenido debido a que los ítems de la escala tienen como mínimo un valor de 0.7, con $p < 0,05$. Con la valoración de los jueces fue excluido un ítem que no cumplió con el valor mínimo establecido y se realizaron algunos ajustes en la redacción a partir de sus observaciones, lo que permitió mejorar el instrumento previo a la aplicación a la muestra de docentes.

3.3. Ejecutar un análisis factorial para valorar sus dimensiones

Primero se indagó si se cumplían las condiciones para el análisis factorial. Al respecto, se encontró que se cuenta con una muestra adecuada, la prueba de esfericidad de Bartlett, así como la prueba KMO indicaron que los datos son adecuados para realizar el análisis factorial. El Análisis Factorial Exploratorio (AFE) mostró la presencia de cuatro factores.

Posteriormente se realizó el Análisis Factorial Confirmatorio (AFC) para verificar el ajuste al modelo y se encontró que cumple con los parámetros establecidos: GFI=0.996; NFI=0.995; RMSEA=0.000.

3.4. Analizar la confiabilidad del instrumento mediante la aplicación del Alfa de Cronbach

La confiabilidad total del instrumento, utilizando el programa estadístico Jasp 14.1, fue de 0.981. En la Tabla 9 se muestran los resultados para cada uno de los cuatro factores que componen el modelo.

Tabla 9

Confiabilidad de los factores de la escala

Factor	Alpha de Cronbach
F1. Asignación y uso de la calificación	0.944
F2. Conocimiento y uso de evaluaciones estandarizadas	0.951
F3. Evaluación formativa	0.944
F4. Uso de la evaluación	0.892

4. Discusión y conclusiones

A partir del estudio realizado puede concluirse que el instrumento para evaluar literacidad evaluativa en docentes de educación básica en México es pertinente para evaluar y diagnosticar si los profesores cuentan con los conocimientos y habilidades necesarios para realizar la evaluación del aprendizaje de sus alumnos.

En la revisión de la literatura no se encuentran instrumentos elaborados específicamente para medir la literacidad evaluativa en docentes de educación básica en México. Aunque sí se ubicaron cuatro referentes que pueden ser un punto de partida para el presente trabajo: 1) La investigación de Arce-Ferrer, A. J., Cab, V. P., y Cisneros-Cohernour, (2001), que consistió de una encuesta aplicada a 169 maestros, consejeros y administradores para valorar sus competencias en evaluación; 2) La investigación de Morán

et al. (2015), que consideró la evaluación de las competencias docentes, en una muestra de 90 profesores, en la asignatura de tecnología en educación secundaria y consideró a la evaluación como una de las variables a medir; 3) Ramón et al., (2017) evaluaron mediante una entrevista a una muestra de 20 docentes de educación media superior en la que incluyeron seis preguntas relacionadas con la evaluación formativa y; 4) La SEP por medio de un documento que guía y orienta la actividad docente en educación básica, establece criterios e indicadores para personal docente y directivo con respecto a la evaluación (SEP, 2019).

Los cuatro antecedentes enunciados, así como la experiencia que estaba experimentando México en la revisión del sistema de rendición de cuentas escolar y la posibilidad de respaldar la toma de decisiones en políticas educativas a partir de resultados derivados de evaluaciones externas (Arce-Ferrer, A. J., Cab, V. P., y Cisneros-Cohernour, 2001), constituían un escenario idóneo para contar con instrumentos pertinentes para valorar en qué medida los docentes estaban realizando lo que se ha establecido en los diferentes documentos de las reformas educativas subsecuentes y, específicamente lo relacionado con la evaluación de los aprendizajes de los alumnos.

El juicio de expertos realizado en este estudio por 11 docentes investigadores con experiencia en el área tuvo como resultado una V de Aiken de la escala total de 0,81 en pertinencia, 0,78 en redacción de los ítems y 0,75 en satisfacción. Al respecto, existen diversos estudios que señalan la importancia del juicio de expertos para asegurar la pertinencia de los instrumentos (Bautista Lima et al., 2019; Escobar-Pérez y Cuervo-Martínez, 2008; Gudiño, 2015; Howell, 2013; Iqbal et al., 2020; Martin-Romera y Ruiz, 2017).

El instrumento elaborado con el propósito de evaluar la literacidad evaluativa en docentes fue sometido al AFE para determinar su estructura y posteriormente, mediante el AFC se validó un modelo con 4 factores. El modelo obtenido en esta investigación no coincidió con los factores que originalmente se encontraron en la revisión de la literatura. Se consideró el trabajo de Howell, (2013), como base para elaborar el instrumento; otra sección se elaboró a partir de la investigación de DeLuca et al., (2016) y, finalmente para los ítems

correspondientes al factor de calificación, se consideró la investigación de Liu, (2008). Lo anterior permitió partir de una estructura que consideró siete factores originalmente.

En relación con la confiabilidad del instrumento, esta fue satisfactoria tanto de manera general (0.981); así como para cada uno de los cuatro factores de la estructura que componen el modelo (F1=0.944, F2=0.951, F3=0.944 y F4=0.892).

El presente estudio contribuye a contar con un instrumento fiable y válido para valorar la literacidad evaluativa en docentes de educación básica en México y recopilar información que permita tomar decisiones pertinentes relacionadas con su formación, capacitación y actualización. Se debe reconocer también la importancia de identificar mediante otras metodologías lo que los docentes conocen y aplican en el aula con respecto a la evaluación del aprendizaje, debido a que el auto reporte es una aproximación y se puede complementar con observaciones en el aula, revisión de los instrumentos de evaluación elaborados por el docente; así como los registros de evaluación y la planificación didáctica.

Referencias

- Alkharusi, H., Aldhafri, S., Alnabhani, H., & Alkalbani, M. (2014). Educational assessment profile of teachers in the Sultanate of Oman. *International Education Studies*, 7(5), 116–137. <https://doi.org/10.5539/ies.v7n5p116>
- Arce-Ferrer, A. J., Cab, V. P., & Cisneros-Cohernour, E. J. (2001). *Teachers' Assessment Competencies*. Annual meeting of the American Educational Research Association (Seattle, WA, April 10-14). <https://files.eric.ed.gov/fulltext/ED455280.pdf>
- Bautista Lima, A. J., Juárez Hernández, L. G., & Tobón, S. (2019). Design and validation of a rubric to evaluate the ethical project of life in university students. *World Review of Science, Technology and Sustainable Development*, 15(4), 300–312. <https://doi.org/10.1504/WRSTSD.2019.104093>
- 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>
- Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language Assessment Literacy: What do we need to learn, unlearn, and relearn? *Language Testing in Asia*, 10.
- Daniel, L. G., & King, D. A. (1998). Knowledge and use of testing and measurement literacy of elementary and secondary teachers. *Journal of Educational Research*, 91(6), 331–344. <https://doi.org/10.1080/00220679809597563>
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: a review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), 251–272. <https://doi.org/10.1007/s11092-015-9233-6>
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Approaches to classroom assessment inventory: A new instrument to support teacher assessment literacy. *Educational Assessment*, 21(4), 248–266. <https://doi.org/10.1080/10627197.2016.1236677>
- Escobar-Pérez, J., & Cuervo-Martínez, Á. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en Medición*, 6, 27–36.
- Pérez García, J., & López Ortega, M. A. (2020). Escala para evaluar la Literacidad Evaluativa en Docentes de Educación Básica. *Transdigital*, 3(6). 1–27. <https://doi.org/10.56162/transdigital159>

- Gudiño, A. (2015). Metodología para determinar las competencias evaluativas de los docentes. *Educación en Contexto*, 1(2), 56–79.
- Guti, L. (2019). Cómo realizar e interpretar un análisis factorial exploratorio utilizando SPSS. *REIRE Revista D'Innovació i Recerca en Educació*, 12(2), 1–14. <https://doi.org/10.1344/reire2019.12.227057>
- Hailaya, W., Alagumalai, S., & Ben, F. (2014). Examining the utility of Assessment Literacy Inventory and its portability to education systems in the Asia Pacific region. *Australian Journal of Education*, 58(3), 297–317. <https://doi.org/10.1177/0004944114542984>
- Howell, C. A. (2013). *Development and analysis of a measurement scale for teacher assessment literacy*. East Carolina University.
- Huang, J., & He, Z. L. (2016). Exploring Assessment Literacy. *Higher Education of Social Science*, 11(2), 18–27. <https://doi.org/10.3968/8727>
- Iqbal, Z., Saleem, K., & Arshad, H. M. (2020). Measuring teachers' knowledge of student assessment: Development and validation of an MCQ test. *Educational Studies*, 49(1), 166–183. <https://doi.org/10.1080/03055698.2020.1835615>
- Khodabakhshzadeh, H., Kafi, Z., & Hosseinnia, M. (2018). Investigating EFL teachers' conceptions and literacy of formative assessment: Constructing and validating an inventory. *International Journal of Instruction*, 11(1), 139. <https://doi.org/10.12973/iji.2018.11110a>
- Liu, X. (2008). Assessing measurement invariance of the teachers' perceptions of grading practices scale across cultures. *NERA Conference Proceedings 2008*. 3, 1(20), 70–80.
- Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2018). Reconceptualising the role of teachers as assessors: teacher assessment identity. *Assessment in Education: Principles, Policy and Practice*, 25(5), 442–467. <https://doi.org/10.1080/0969594X.2016.1268090>
- Martin-Romera, A., & Ruiz, E. M. (2017). Valor del conocimiento pedagógico para la docencia en Educación Secundaria: Diseño y validación de un cuestionario. *Estudios Pedagógicos*, 43(2), 195–220. <https://doi.org/10.4067/S0718-07052017000200011>
- Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of
- Pérez García, J., & López Ortega, M. A. (2020). Escala para evaluar la Literacidad Evaluativa en Docentes de Educación Básica. *Transdigital*, 3(6). 1–27. <https://doi.org/10.56162/transdigital159>

classroom assessment professional development. *Improving Schools*, 12(2), 101–113.
<https://doi.org/10.1177/1365480209105575>

Mertler, C. A., & Campbell, C. (2005). Measuring teachers' knowledge & application of classroom assessment concepts: Development of the assessment literacy inventory. *American Educational Research Association*, 27.

Montero, I., & León, O. G. (2007). A guide for naming research studies in Psychology. *International Journal of Clinical and Health Psychology*, 7(3), 847–862.

Morán, R., Cardoso, E. O., Cerecedo, M. T., & Ortíz, J. C. (2015). Evaluación de las competencias docentes de profesores formados en instituciones de educación superior: El caso de la asignatura de tecnología en la enseñanza secundaria. *Formacion Universitaria*, 8(3), 57–64.
<https://doi.org/10.4067/S0718-50062015000300007>

Morata, M. Á., Holgado, F. P., Barbero, M. I., & Mendez, G. (2015). Análisis factorial confirmatorio. Recomendaciones sobre mínimos cuadrados no ponderados en función del error Tipo I de Ji-Cuadrado y RMSEA. *Acción Psicológica*, 12(1), 79.
<http://revistas.uned.es/index.php/accionpsicologica/article/view/14362>

Nikmard, F., & Mohamadi Zenouzagh, Z. (2020). Designing and validating a potential assessment inventory for assessing ELTs' assessment literacy. *Language Testing in Asia*, 10(1).
<https://doi.org/10.1186/s40468-020-00106-1>

OECD. (2019). *PISA 2018 assessment and analytical framework*. PISA, OECD Publishing.
<https://doi.org/10.1787/b25efab8-en>

Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, 12(4), 10–12.
<https://doi.org/10.1111/j.1745-3992.1993.tb00548.x>

Ramón, P., Aquino, S., & Alejandro, M. (2017). Evaluación de las competencias docentes de profesores de Educación media superior del Estado de Tabasco, México. *Sinéctica*, 48, 18.

SEP. (2019). Perfiles profesionales, criterios e indicadores para docentes, técnicos docentes y personal con funciones de dirección y de supervisión. *DOF 2019*, 62.

Stiggins, R. (2017). *The perfect assessment system*. ASCD.

Tagele, A., & Bedilu, L. (2015). Teachers competence in the educational assessment of students: The case of secondary school teachers in the Amhara National Regional State. *The Ethiopian Journal of Education*, 35(2), 163–191.

Yamtim, V., & Wongwanich, S. (2014). A study of classroom assessment literacy of primary school teachers. *Procedia - Social and Behavioral Sciences*, 116, 2998–3004. <https://doi.org/10.1016/j.sbspro.2014.01.696>

Zhang, Z., & Burry-Stock, J. (1997). Assessment Practices Inventory: A multivariate analysis of teachers' perceived assessment competency. *Annual Meeting of the National Council on Measurement in Education*, Chicago, IL.