

Evaluación del lenguaje en preescolar: desarrollo y validación de una rúbrica

Language assessment in kindergarten: Development and validation of a rubric

SEIRI ADILENE GARCÍA ALDACO*

LUIS HORACIO PEDROZA ZÚÑIGA**

El artículo presenta el resultado de una investigación cuyo objetivo fue el desarrollo y la obtención de evidencias de validez de un instrumento de evaluación de lenguaje para alumnos de preescolar. El estudio metodológico de evaluación educativa se dividió en dos fases: el diseño del instrumento, basado en las directrices de estándares internacionales para la elaboración de instrumentos de medición, y el logro de evidencias de validez de contenido a través del juicio de expertos y evidencias de validez fundamentadas en la estructura interna mediante análisis factorial exploratorio. Se obtuvo una rúbrica alineada al currículo y administrada por las educadoras que permite valorar el desempeño de niños y niñas, tanto en lenguaje oral como escrito, para fines formativos y de monitoreo de tendencias. La rúbrica mostró alta confiabilidad y una estructura unidimensional, y se recabaron evidencias de validez que respaldan su contenido. Las características del instrumento lo hacen valioso para su uso en la educación preescolar mexicana.

The objective of this research was to develop and obtain validity evidence of a language assessment instrument for kindergarten students. A methodological study of educational evaluation was conducted, it was divided into two phases: Instrument development, which was based on the guidelines of international standards for the development of measurement instruments; and a second phase dedicated to obtaining evidence based on test content through expert judgment, and evidence based on internal structure applying exploratory factor analysis. The rubric obtained is aligned to the curriculum, administered by teachers, appropriate to assess children's performance, in both oral and written language, and can be used for formative purposes and monitoring trends. The rubric's content is supported by evidence, and the results indicated the rubric has high reliability and a one-dimensional structure. The characteristics of this instrument, which make evident that it is adequate to be used in the Mexican kindergarten education system, are discussed.

Palabras clave:
lenguaje,
educación
preescolar,
instrumentos de
evaluación,
evaluación del
aprendizaje,
evaluación de
estudiantes

Keywords:
language
acquisition,
kindergarten
children,
curriculum-based
assessment,
criterion referenced
tests,
performance-based
assessment

Recibido: 13 de octubre de 2021 | **Aceptado para su publicación:** 26 de abril de 2022 | **Publicado:** 3 de mayo de 2022

Cómo citar: García Aldaco, S. A. y Pedroza Zúñiga, L. H. (2022). Evaluación del lenguaje en preescolar: desarrollo y validación de una rúbrica. *Sinéctica, Revista Electrónica de Educación*, (58), e1375. [https://doi.org/10.31391/S2007-7033\(2022\)0058-015](https://doi.org/10.31391/S2007-7033(2022)0058-015)

* Maestra en Ciencias Educativas. Auxiliar investigador en el Instituto de Investigación y Desarrollo Educativo de la Universidad Autónoma de Baja California. Línea de investigación: evaluación educativa. Correo electrónico: seiri.garcia@uabc.edu.mx/<https://orcid.org/0000-0002-6262-5650>

** Doctorado en Ciencias Educativa. Investigador en el Instituto de Investigación y Desarrollo Educativo de la Universidad Autónoma de Baja California. Línea de investigación: evaluación educativa, investigación de la educación preescolar, y evaluación formativa de docentes y alumnos. Correo electrónico: horacio.pedroza@uabc.edu.mx/<https://orcid.org/0000-0002-5256-2967>



Introducción

El aprendizaje de la lengua es uno de los propósitos más importantes de los sistemas educativos. Lograr que las personas puedan comunicarse de forma adecuada tanto en materia oral como escrita resulta fundamental para la niñez en su etapa de escolarización, así como para su posterior inserción en la vida social. En México, en todos los niveles de educación básica, el campo de formación académica de lenguaje y comunicación se clasifica en cinco asignaturas (Lengua materna-español, Lengua materna-lengua indígena, Segunda lengua-lengua indígena, Segunda lengua-español y Lengua extranjera-inglés), que comparten la visión general de lenguaje; en especial, la asignatura Lengua materna-español se orienta a la promoción de la lengua en español que el alumno adquiere de los adultos que lo rodean en su contexto natural (Secretaría de Educación Pública [SEP], 2017). En el nivel preescolar, por las características de los alumnos, el principal enfoque de este campo es la promoción de la oralidad y la integración a la cultura escrita (SEP, 2017).

De acuerdo con la SEP (2017), que es la dependencia del Poder Ejecutivo federal encargada de la administración, regulación y fomento de la educación en México, la importancia de la adquisición de los contenidos del campo de lenguaje y comunicación radica en dos vertientes. Primero, la oralidad es mencionada como el preámbulo para adquirir la lengua escrita, mientras que el campo de formación en general es considerado como la base para el logro de otros aprendizajes. Segundo, el lenguaje incide en el desarrollo cognitivo y emocional, ya que es una herramienta para la construcción de conocimientos y significados que le permiten al alumno integrarse a su sociedad y cultura, así como adquirir seguridad y confianza en sí mismo.

En este punto, las evaluaciones realizadas a nivel nacional en 2007 y 2011 por el Examen de la Calidad y el Logro Educativo para preescolar (EXCALE 00) revelaron que el mayor porcentaje de alumnos de tercer grado de preescolar se encontraba en el nivel básico en lenguaje (niveles de desempeño de la prueba: por debajo del básico, básico, medio y avanzado). En este nivel de dominio, los alumnos solamente son capaces de reconocer su nombre escrito y mencionarlo, proporcionar información básica de su familia, intercambiar opiniones de acuerdo o desacuerdo sobre un tema y seguir instrucciones (Instituto Nacional de Evaluación Educativa [INEE], 2014).

Ahora bien, la evaluación de alumnos en edad preescolar (tres a cinco años) tiene características particulares que la hacen en especial difícil; entre ellas destacan tres: la primera es que las estrategias de evaluación deben coincidir con el nivel de desarrollo físico, mental y social del niño (Wortham & Hardin, 2016); la selección inadecuada de estrategias conlleva a la obtención de datos insuficientes que limitan la evaluación y, con ello, se compromete la confiabilidad de esta (Snow & Van-Hemel, 2008). La segunda es que en esa edad los alumnos aún no leen o escriben de forma convencional, lo que dificulta utilizar evaluaciones a lápiz y papel (Shepard et al., 1998; Wortham & Hardin, 2016). La tercera característica es que los niños necesitan contextos familiares para poder demostrar sus habilidades (Shepard et al., 1998), ya que la evaluación por agentes externos, sin familiaridad con los alumnos, puede resultar amenazante para ellos e interferir en su desempeño (Moreno, 2015).

A partir de los usos inapropiados de las evaluaciones, se establecieron algunos principios para el desarrollo y uso de estas, los cuales señalan que la evaluación a niños y niñas debería generar beneficios para ellos mismos; adaptarse a un propósito específico y ser

confiables, válidas y justas para ese objetivo; reconocer que la confiabilidad y la validez de las evaluaciones aumentan con la edad de los niños; ser apropiadas para la edad tanto en contenido como en método de recolección de datos; y los padres y docentes son una fuente de información valiosa, además de una audiencia para los resultados de la evaluación (Shepard et al., 1998).

La tabla 1 incluye tres de los propósitos de las evaluaciones en niños y niñas de tres a cinco años de edad, así como los usos pretendidos y requisitos técnicos que le corresponden a cada uno. El primero se refiere a una evaluación para el aprendizaje; el segundo es para identificar servicios de salud y necesidades especiales, mientras que el tercero es para reconocer tendencias y evaluación de los programas. Cabe señalar que estos propósitos podrían conjugarse, especialmente el primero y el tercero, siempre y cuando se adopten los requisitos técnicos necesarios (Shepard et al., 1998).

Tabla 1. Propósitos, usos y características técnicas de las evaluaciones en niños y niñas de tres a cinco años

Propósito	Usos	Requisitos técnicos
1. Evaluación para promover el aprendizaje y el desarrollo de los niños	Los padres, cuidadores y docentes de preescolar utilizan medidas directas, incluidas observaciones de lo que los niños están aprendiendo, para decidir qué enseñar a continuación	Los requisitos de confiabilidad y validez son los menos exigentes de todos los propósitos. Pueden ser evaluaciones informales durante la enseñanza Las evaluaciones más formales, realizadas para mejorar el aprendizaje, deben estar vinculadas al plan de estudios de preescolar y tener claras implicaciones sobre qué hacer a continuación. Para ello, los docentes deben conocer la progresión típica de las competencias de los niños y estar familiarizados con las expectativas de edad y grado para reconocer desempeños precoces, por debajo o de acuerdo con lo esperado
2. Identificar niños para servicios de salud y educación especial	Los niños que ingresan a programas preescolares deben someterse a pruebas de identificación de necesidades de salud, incluidos controles de vista y audición. Los niños con posibles retrasos en el desarrollo deben ser referidos para una evaluación en profundidad	Los instrumentos se aplican en dos pasos: una primera aplicación con instrumentos y examinadores menos exactos para una posible identificación; una segunda, en caso de detección, realizada con instrumentos más confiables y especialistas en el área a evaluar
3. Monitoreo de tendencias y evaluación de programas y servicios	Las evaluaciones, incluidas las medidas directas e indirectas del desarrollo de los niños, podrían construirse y usarse para evaluar los programas de preescolar, pero tales medidas no serían lo suficientemente precisas para tomar decisiones importantes sobre niños individuales	Deben seguir altos estándares técnicos para asegurar la precisión Ser ampliamente representativas de las dimensiones del aprendizaje y el desarrollo infantil temprano Pueden aplicarse con diseño matricial para evitar procesos extenuantes para los niños. Sin embargo, no se obtienen resultados por niño, pero permiten conocer tendencias a gran escala de la idoneidad de los programas

Fuente. Shepard et al. (1998)/American Psychological Association-Coalition for Psychology in Schools and Education (2019).

Con base en el panorama planteado, el análisis sobre el desarrollo de instrumentos de evaluación del lenguaje en el ámbito internacional arroja un número considerable de herramientas referidas a la norma, en su mayoría diseñadas en Estados Unidos y Canadá (Fernald et al., 2017). En distintos países de Latinoamérica se han utilizado

instrumentos en principio diseñados para angloparlantes, tal es el caso del Early Grade Reading Assessment, concebido para establecer una línea base de las habilidades para la adquisición de lectura, así como evaluar programas educativos a gran escala, lo que es coincidente con los propósitos uno y tres de la tabla 1. Mide aspectos como la identificación del nombre de la letra, identificación sonoro-gráfica, decodificación, segmentación silábica, entre otros (Dubeck & Gove, 2015). Es evidente que los aspectos medidos por este instrumento no son concordantes con el enfoque socioconstructivista que prevalece en los currículos de la mayoría de países latinoamericanos.

Otro elemento ampliamente utilizado en los países de Latinoamérica es la evaluación infantil temprana (The Learning Bar, 2021), cuyo objetivo es valorar el desarrollo infantil en cinco dominios, incluido el lenguaje –propósitos uno y tres–. A partir de un conjunto de rúbricas, los docentes evalúan a los infantes en su desempeño cotidiano en el aula (The Learning Bar, 2021; Willms, 2014). Esta evaluación se ha utilizado de manera censal en países como Uruguay, con aplicaciones al inicio y al final del ciclo escolar; se ha podido reconocer a niños en riesgo y realizar intervención temprana en el ciclo escolar (López & Willms, 2020).

En México, el Instituto Nacional para la Evaluación de la Educación (INEE) puso en práctica el Examen de Calidad y Logro Educativo para preescolar (EXCALE 00) con el objetivo de proveer información sobre el aprendizaje del currículo a un nivel de agregación nacional –propósito tres– y por tipo de servicio educativo. Esta prueba estandarizada es aplicable por un evaluador externo y, dado su carácter matricial, no es posible obtener resultados individuales (INEE, 2014).

Marín, Guzmán y Castro (2012) elaboraron un instrumento para evaluar competencias de los niños en edad preescolar sobre el área de lenguaje y teniendo como referente el Programa de Educación Preescolar de 2004 (PEP 04). Con la finalidad de capturar el desempeño de los niños en tareas familiares, los autores diseñaron tareas evaluativas que pueden aplicarse en tres formatos: por la educadora, por observador externo con apoyo de la educadora, y en circuitos o estaciones –propósitos uno y tres–. El instrumento mostró adecuados índices de confiabilidad en cada una de las modalidades probadas y se aportaron evidencias de validez del contenido.

Otro instrumento utilizado es el *Manual de exploración de habilidades básicas en escritura, lectura y conteo*, un examen estructurado aplicable por la educadora de manera estandarizada, con el objetivo de evaluar componentes específicos del proceso de alfabetización inicial (SEP, 2018) –propósitos uno y tres–. Este manual presenta las siguientes problemáticas: no brinda un seguimiento del avance en el dominio de los contenidos a lo largo de los tres años de preescolar y, por tanto, se cuenta con un menor tiempo para llevar a cabo estrategias remediales respecto a contenidos no desarrollados de la forma esperada; no muestra evidencias de validez; únicamente evalúa contenidos de lectura y escritura, y deja de lado aspectos de la oralidad, aun cuando esta es una prioridad en educación preescolar (SEP, 2017).

Sobre las evaluaciones que se enfocan solo a uno de los aspectos del lenguaje, por ejemplo, la lectura o escritura, tenemos que reconocer que ello implica tener una subrepresentación del constructo, además de posibles consecuencias sociales no deseadas por la evaluación. Black y William (2018) señalan que este es un problema común de las evaluaciones: centrarse en lo que puede evaluarse de forma estructurada, pero también tiene consecuencias sociales no deseadas, pues origina que los

docentes se centren en evaluar apenas algunos aspectos o en enseñar los contenidos evaluados, y omiten elementos importantes del aprendizaje.

Los diseñadores de instrumentos han implementado algunas medidas que les ayuden a enfrentar las dificultades de la evaluación de los niños y niñas de edad preescolar. Entre estas, podemos destacar la evaluación a partir de desempeños y tareas auténticas, el uso de rúbricas y el involucramiento de las educadoras para realizar las valoraciones (Marín, Guzmán y Castro, 2012; Shepard et al., 1998). Estas características les han permitido tener propiedades métricas robustas y obtener resultados por alumno en evaluaciones a gran escala e, incluso, ofrecer información para ser utilizada con fines de mejora por las educadoras (López & Willms, 2020).

Una de las tareas más complejas de los docentes es la evaluación, por lo que no es raro que se reconozca como un área de mejora (García et al., 2011; Martínez-Rizo, 2012a, 2012b). En educación preescolar, las educadoras también manifiestan dificultades en esas prácticas, como problemas al evaluar desde el enfoque de competencias (Marín, Guzmán y Castro, 2012), el uso inadecuado de técnicas como la observación y la falta de formación para realizar evaluaciones (Gómez et al., 2018).

No obstante, la evaluación en preescolar tiene la posibilidad de promover la mejora en el desempeño de los alumnos. En el aula, es capaz de favorecer aspectos referentes al aprendizaje y enseñanza a partir de identificar las metas, reconocer el estado en el cual se encuentran los niños y las niñas, y promover la retroalimentación para alcanzar niveles ulteriores de desarrollo, además de fomentar procesos metacognitivos y de autorregulación (Black & William, 2019; Hattie & Timperley, 2007). Al docente le permite reflexionar y decidir sobre el curso posterior de su práctica (Martínez, 2012a, 2012b). Desde una visión macro, la información que proporciona es también valiosa, pues permite al sistema educativo reconocer los avances o rezagos que se tienen y tomar decisiones informadas.

La investigación sobre la evaluación de aprendizajes en educación preescolar se encuentra en un estado incipiente (Gómez et al., 2018). Queda de manifiesto la necesidad de contar con instrumentos de evaluación que faciliten valorar las competencias del currículo (Cardemil y Roman, 2014; Gómez et al., 2018; Gómez y Seda, 2008; Marín, Guzmán y Castro, 2012) y sean sensibles a las características y el desarrollo de los infantes en edad preescolar.

Con base en lo mencionado, el propósito de este artículo es presentar el desarrollo y la obtención de evidencias de validez de un instrumento de evaluación de lenguaje en español como lengua materna para alumnos de preescolar alineado con el currículo en el contexto mexicano.

MÉTODO

Para la consecución del objetivo, llevamos a cabo un estudio metodológico de evaluación educativa; de acuerdo con Jornet, González y Suárez (2010) y Jornet, Perales y González (2020), una rama de este tipo de estudios se orienta a la construcción y validación de instrumentos de evaluación.

Participantes

Tuvimos tres tipos de participantes: comités, educadoras frente a grupo y alumnos de educación preescolar. Formamos tres comités de jueces:

- Comité de jueceo de diseño. Compuesto por dos expertos con amplia experiencia en el nivel preescolar: uno en investigación y evaluación educativa y otro con conocimiento del currículo y las prácticas educativas.
- Comité de indicadores de desempeño. Formado por cuatro expertos, tres en lenguaje y comunicación para preescolar y uno en el área de desarrollo de instrumentos de medición.
- Comité de jueceo del instrumento. Integrado por cuatro expertos en lenguaje y comunicación para preescolar y con experiencia en la práctica docente.

Tabla 2. Experiencia de los jueces participantes y comités en los que colaboraron

Juez	Puesto actual	Experiencia en educación preescolar o en área de lenguaje	Años de experiencia	Comité en que participó		
				Diseño	Indicadores	Jueceo
1	Jefa de sector de preescolar	Desarrollo del currículo de preescolar a nivel nacional. Evaluación a gran escala de prácticas docentes. Formación de docentes en servicio. Jefa de sector. Docente de escuela normal. Educadora frente a grupo	25 años	X		X
2	Investigador	Investigación en el nivel preescolar. Evaluación a gran escala de alumnos, docentes y escuelas. Diseño de instrumentos de evaluación. Estimulación del desarrollo infantil	17	X	X	
3	Profesora investigadora	Investigación en enseñanza de la lengua en preescolar y primaria. Docente universitaria de licenciatura y posgrado en área de lenguaje	13 años		X	X
4	ATP	Asesora técnico pedagógica. Docente de escuela normal. Investigación en enseñanza de la lengua en edad preescolar. Educadora frente a grupo	6 años		X	X
5	Estudiante de doctorado	Investigación en enseñanza de la lengua en edad preescolar y en escuelas multigrado. Diseño de programas para el desarrollo de la lengua escrita en educación indígena. Docente universitaria de licenciatura y posgrado en área de lenguaje	8 años		X	X

Además, participaron 11 educadoras frente a grupo y 241 niños y niñas (105 y 136, respectivamente) de educación preescolar de los grados segundo y tercero (111 y 130, en ese orden). Normativamente, los alumnos de segundo de preescolar deben tener cuatro años cumplidos y los de tercero, cinco. Tanto educadoras que fungieron como aplicadoras como los niños a los que se les suministró la rúbrica fueron seleccionados por conveniencia.

Procedimiento

El procedimiento se conformó de dos grandes etapas que se desarrollaron en forma paralela. La primera estuvo dedicada a la construcción del instrumento de evaluación; y la segunda, orientada a la obtención de evidencias de validez basadas en el contenido y la estructura interna.

Etapa 1. Desarrollo del instrumento de evaluación

En la figura observamos que, para la consecución de la primera etapa del estudio, hicimos una adaptación de la propuesta de Lane y colaboradores (2016) para la construcción de instrumentos de evaluación, la cual considera los criterios de calidad de los estándares de la American Educational Research Association/American Psychological Association/National Council on Measurement in Education (AERA/APA/NCME) (2014).

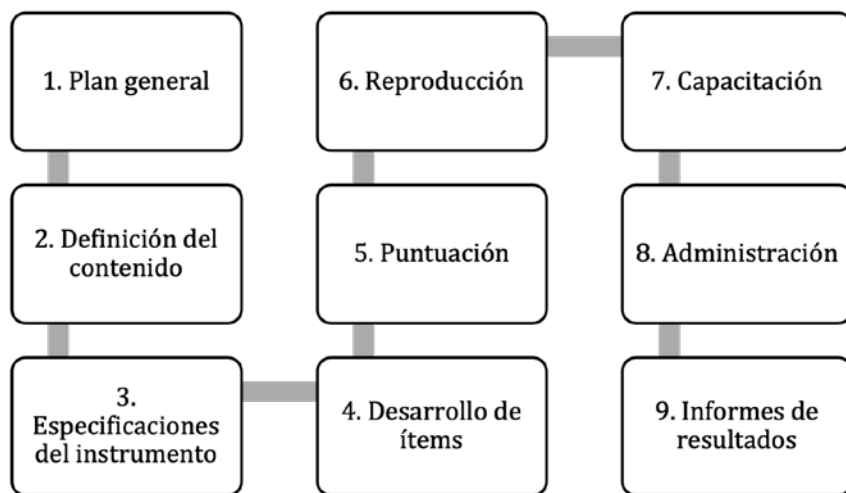


Figura. Procedimiento para el desarrollo de la rúbrica.
Nota: Adaptado del *test development process* (Lane et al., 2016).

Plan general. Este componente describe los aspectos relevantes de la construcción del instrumento: contenido de evaluación, propósito, tipo de instrumento, referente (norma o criterio), clase de reactivos, usuario, forma de administración, entre otros (Lane et al., 2016).

Definición del contenido. Las actividades de este componente se relacionan con la delimitación del contenido a evaluar y las afirmaciones que se harán a partir del instrumento (Lane et al., 2016).

Para este instrumento, llevamos a cabo un análisis del currículo nacional para preescolar, a partir del cual construimos una estructura inicial valorada por el comité de jueceo de diseño. Posteriormente, desarrollamos una segunda versión de estructura, seleccionamos los contenidos a evaluar y los clasificamos dentro de la estructura prevista; el producto final fue presentado al comité de indicadores de desempeño, el cual agregó recomendaciones para la mejora y contribuyó al planteamiento de los indicadores y sus descriptores, además de ejemplos de desempeño. Por último, la información obtenida se trianguló con lo establecido en el currículo y la teoría.

Especificaciones del instrumento. De acuerdo con la AERA/APA/NCME (2014), este tipo de documento articula el alcance del dominio, los contenidos que se medirán, los procesos cognitivos requeridos, los formatos de los ítems y la clase de respuesta.

Las especificaciones de los ítems para el instrumento construido contienen descripción del contenido a evaluar, los contenidos que abarca cada aspecto de la rúbrica, indicadores graduados por nivel de desempeño y descripciones del formato de la rúbrica.

En cuanto al alcance del dominio, tomamos como base los aprendizajes esperados, así como las teorías con base en las que se plantea el currículo. En total, formulamos siete especificaciones referentes a cada uno de los ítems de la rúbrica.

Los indicadores de desempeño se definieron de manera conjunta con el comité del mismo nombre, y seguimos el procedimiento empleado por Pedroza y Luna (2017). A partir de una propuesta de indicadores de desempeño para cada ítem, en conjunto, los autores y los miembros del comité establecieron niveles de desempeño para cada indicador, así como ejemplos prototípicos. El trabajo se realizó en una sesión plenaria de seis horas y luego en interacciones individuales, presenciales y asincrónicas.

Desarrollo de ítems. Este paso corresponde a la construcción de los ítems (Lane et al., 2016); acorde con la lógica del instrumento elaborado, este paso se dedicó a la elaboración de las descripciones de los niveles de desempeño de cada aspecto de la rúbrica. En seguida, el comité de jueceo del instrumento se encargó de su valoración.

Puntuación. Las actividades en este componente se enfocan en el diseño de procedimientos para garantizar la integridad de todo el proceso de asignación de niveles de desempeño; para ello, se requieren pautas de puntaje bien elaboradas y evaluadores capacitados (Lane et al., 2016).

Para el método con que se puntúa la rúbrica, hicimos un acercamiento a docentes de preescolar. A partir de sus aportaciones, determinamos una lógica de puntuación para establecer la consolidación de los niveles de desempeño que plantea la teoría del desarrollo del lenguaje, una lógica para los alumnos cuyo dominio estuviera en desarrollo con base en la función de andamiaje de las educadoras –postulada en la teoría socioconstructivista y el currículo (SEP, 2017)– y una lógica para los alumnos que hubieran excedido los aprendizajes esperados para preescolar, con fundamento en los criterios estipulados en los programas de estudios para ese nivel.

Reproducción. El proceso de este componente se centra en la selección del formato o medio para reproducir el instrumento de evaluación, así como la confección de los elementos que habrán de acompañarlo, como formatos o materiales auxiliares (Lane et al., 2016).

Los elementos constitutivos del instrumento son: rúbrica, manual de aplicación y formato de registro. Para nuestro estudio, decidimos utilizar una versión impresa y una digital en línea; esta última facilitó el resguardo y acceso a los datos, el seguimiento de las evaluaciones y la producción de reportes generales por alumno y grupo. Las educadoras participantes tuvieron acceso a todos los materiales y a ambas versiones.

Capacitación. Llevamos a cabo actividades relacionadas con el contacto de las personas que dieron acceso a las instituciones donde se efectuaron las evaluaciones, la estandarización de la administración y aspectos éticos.

Una acción fundamental fue la implementación de un taller de capacitación para el uso de del instrumento. Este se diseñó para guiar al profesorado respecto al contenido de evaluación y los criterios de los niveles de desempeño de la rúbrica. Aunado a lo anterior, llevamos a cabo un ejercicio de calibración, que consistió en presentar siete videos de alumnos de preescolar en distintas situaciones pedagógicas –previamente analizados y puntuados por los desarrolladores del instrumento– y solicitar a las educadoras que ubicaran el nivel de desempeño de cada niño. El porcentaje de acuerdo exacto fue del 76% y el porcentaje de acuerdo entre categorías adyacentes (+/- 1) fue del 21%. Además, capacitamos a las docentes en el uso del instrumento en línea para el registro de la información.

Administración. Este componente abarca las actividades de la administración del instrumento (Lane et al., 2016); dado su impacto en la validez de las interpretaciones y usos de los puntajes, AERA/APA/NCME (2014) hacen hincapié en minimizar errores en la aplicación.

La administración del instrumento se realizó en tres semanas, agendado acorde con el periodo que el sistema educativo normalmente ofrece para la evaluación de los niños. A partir de las actividades que la educadora diseña para la evaluación, se utiliza el instrumento para hacer la valoración de grupos de alumnos a través de la observación de sus actividades. El proceso de administración es similar al de otros instrumentos, como la evaluación infantil temprana (López & Willms, 2020).

Informes de resultados. El producto de este componente es uno de los más destacados de los instrumentos, pues sirve como base para inferencias, decisiones y acciones (Lane et al., 2016). Para este estudio, elaboramos reportes individuales por alumno mediante la versión en línea del instrumento dirigidos al profesorado de preescolar.

Etapa 2. Obtención de evidencias de validez

El procedimiento para la obtención de evidencias de validez de los productos de la etapa 2 se muestra en la tabla 3. Para efectos de claridad, este apartado está estructurado de acuerdo con la fuente de evidencia de validez recabada.

Tabla 3. Procedimiento de la etapa 2

Fases	Aspectos analizados	Evidencias de validez
1. Evidencias de validez de contenido	Estructura del instrumento, tabla de selección y clasificación de contenidos	Comité de jueceo de diseño
		Comité de indicadores
	Operacionalización de las variables	Comité de jueceo del instrumento
	Versión previa del instrumento	
Puntuaciones obtenidas del jueceo del instrumento	Coefficiente de CVC*	
2. Evidencias de validez de estructura interna	Dimensionalidad y confiabilidad del instrumento	AFE**
		Alfa de Cronbach

*Coeficiente de validez de contenido

** Análisis factorial exploratorio

Fase 1. Evidencias de validez basadas en el contenido

Para este cometido, utilizamos el juicio de expertos (AERA/APA/NCME, 2014; Sireci & Faulkner, 2014). Tres comités valoraron el contenido del instrumento en concordancia con lo presentado en la tabla 3. En especial, el comité de jueceo del instrumento hizo una valoración individual mediante un cuestionario compuesto de una escala de cinco niveles, en la que se valoró la concordancia de cada ítem y los niveles de desempeño conforme a los siguientes criterios:

- Relevancia. La información contenida es lo suficientemente importante para ser incluida en el instrumento.
- Congruencia. Existe una relación lógica de la información con lo que se evalúa.
- Claridad. El contenido evoca un mismo rasgo y no es confuso.
- Suficiencia. Están incluidos los indicadores y criterios necesarios para evaluar cada dimensión.

Los datos fueron sistematizados a través del coeficiente de validez de contenido total (CVC_t) (Hernández-Nieto, 2011), cuya ecuación se basa en las medias para calcular el acuerdo entre expertos y la validez de contenido. Con base en este autor, se recomienda mantener los ítems con un CVC superior a .80, siguiendo la ecuación (Hernández-Nieto, 2002, p. 72):

$$CVC_t = \frac{\sum Cvc_{tc}}{N} = \sum \left[\left[\frac{\sum x_i}{J} \right] - p_{ei} \right] \left(\frac{1}{N} \right)$$

Donde:

N = número total de ítems del instrumento

$\sum x_i$ = sumatoria de los puntajes asignados por cada juez (J) a cada uno de los ítems (i) (sumatoria de todos los promedios de todos los criterios a evaluar)

Vmx = valor máximo de la escala utilizada por los jueces

P_{ei} = probabilidad de error por cada ítem o de concordancia aleatoria entre jueces

J = número de jueces asignados al ítem

Fase 2. Evidencias de validez basadas en la estructura interna

Utilizamos el análisis factorial exploratorio (AFE) para examinar el conjunto de variables latentes o factores comunes que explican las respuestas a los ítems del instrumento de evaluación y el grado en que las categorías se mezclan a lo

largo de estas dimensiones (Lloret-Segura et al., 2014). Como paso previo, los supuestos requeridos para el uso del AFE se cumplieron mediante la aplicación del índice Kaiser, Meyer y Olkin (KMO) y la prueba de esfericidad de Bartlett. Asimismo, dado que la métrica de los ítems es una escala de siete niveles, los datos se trataron como variables pseudocontinuas, pues, de acuerdo con Bentler (2006), en este caso las distorsiones presentadas en los análisis son mínimas. Además, la consistencia interna del instrumento se determinó con el coeficiente de confiabilidad alfa de Cronbach. El software utilizado para llevar a cabo los análisis correspondientes fue el SPSS Statistics 22.

RESULTADOS

Instrumento desarrollado: rúbrica de evaluación formativa de lenguaje con alineación curricular para preescolar

Esta rúbrica es un instrumento para evaluar el desempeño de niños y niñas de preescolar en el campo de lenguaje y comunicación de la lengua materna español, referido al currículo con fines formativos y monitoreo de tendencias. La educadora frente a grupo es quien valora el desempeño del instrumento con las pautas del manual de aplicación. El instrumento puede utilizarse con un formato en papel o digital y genera reportes individuales en términos descriptivos. Como apreciamos en la tabla 4, se compone de dos ámbitos, el lenguaje oral y el escrito, y cada uno se subdivide en producción e interpretación de la lengua; en total se compone de siete ítems basados en indicadores teóricos y prácticos. Como es notorio, hay una prevalencia del lenguaje oral sobre el escrito debido a que en el currículo la oralidad es el aspecto más relevante en la educación preescolar (SEP, 2017).

Tabla 4. Estructura de la rúbrica e indicadores

Ámbito	Subámbito	Ítems	Indicadores
Lengua oral	Producción	1. Conversación	Concreción y congruencia Aporte de información Gramática, semántica y organización de información
		2. Descripción	Enlista características, congruencia, detalla y discrimina
		3. Narración	Sigue secuencias, ordena ideas, sintetiza, une información y ordena temporalmente Narra, detalla información, selecciona información, coherencia, congruencia, inventa y describe Volumen, entonación, adecuación del lenguaje, estilo del habla, recursos literarios
		4. Explicación	Comprende lo que se le solicita, expresa, argumenta, verbaliza procedimientos, responde, determina su postura, anticipa y responde preguntas Organiza hechos e ideas, reconstruye secuencias y planea orden o pasos Uso de material de apoyo, vocabulario específico y uso de lenguaje especializado
	Interpretación	5. Interpretación de la lengua oral	Competencia comunicativa o pragmática Semántica, pronunciación y gramática

Ámbito	Subámbito	Ítems	Indicadores
Lengua escrita	Producción	6. Producción de la lengua escrita	Patrones evolutivos de la escritura, convenciones de la lengua escrita, escritura silábica, variedad intra-figural: cantidad mínima y no repetir letras, uso de grafemas no icónicos (verdaderas letras, pseudoletas, cuasiletas o números) y estrategias para resolver los problemas de la comprensión del sistema alfabético
	Interpretación	7. Interpretación de la lengua escrita	Procesamiento de texto como lenguaje y decodificar fonema-graía Construcción de significado, proximidad, contexto interno, contexto externo, cantidad mínima y modular

Los ítems de la rúbrica tienen siete niveles de desempeño, tres de ellos con descripciones sintéticas del desempeño. La tabla 5 muestra un ejemplo del ítem referente a conversación. El nivel uno se utiliza para ubicar el desempeño de un alumno cuando no cumple por completo con los criterios establecidos en el nivel dos. Los niveles pares dos, cuatro y seis contienen descriptores del nivel de desempeño de forma ascendente. Se emplean para ubicar el desempeño del alumno solo si cumple por sí mismo y sin apoyo con todos los criterios determinados en el nivel. Los niveles impares tres y cinco sirven para ubicar el nivel de desempeño del alumno cuando ha superado los descriptores del nivel anterior; pero no cumple por completo con los descriptores del nivel sucesivo, o cuando los cubre totalmente solo con apoyo del docente.

El nivel siete se utiliza para ubicar el desempeño del alumno cuando se considera que excede los indicadores propuestos en el nivel anterior (seis) o demuestra un desempeño superior; por ejemplo, que en producción de la lengua escrita el alumno ya escriba de forma convencional por sí mismo, desempeño que en el currículo se propone hasta la educación primaria.

Tabla 5. Ejemplo de la rúbrica referente al ítem de conversación

Nivel 1	No alcanza todos los indicadores del nivel 2
Nivel 2	Se limita a comunicar lo que quiere o se le pregunta con afirmaciones o negaciones sin aportar información nueva. Por su forma de comunicarse, hay ausencia de la organización de la información del discurso, uso de las normas que regulan el lenguaje y adaptación a la audiencia. Ejemplo: "sí", "no" o usa palabras sueltas como "baño", "agua"
Nivel 3	Alcanza algunos indicadores del nivel 4 o todos solo con apoyo de la docente
Nivel 4	Logra comunicar lo que quiere, pero aún no domina del todo la organización de la información de su discurso, el uso de las normas convencionales que estructuran el lenguaje y la adaptación del mensaje a la audiencia. Puede ser redundante en sus aportaciones. Ejemplo: usa frases más elaboradas como "ir baño", "vaso con agua", "gato bonito", "amo mi mami"
Nivel 5	Alcanza algunos indicadores del nivel 6 o todos solo con apoyo de la docente
Nivel 6	Logra comunicar lo que quiere organizando la información de su discurso, haciendo uso de las normas convencionales que estructuran el lenguaje y adecuando el mensaje a la audiencia. Además, aporta información nueva relacionada o acorde con el objetivo o situación comunicativa. Ejemplo: "quiero un vaso de agua", "me das permiso de ir al baño", "ayer fui al parque"
Nivel 7	Supera los indicadores del nivel 6

Las descripciones sintéticas se acompañaron de una descripción en extenso (ver tabla 6), incluidas en el manual de aplicación, el cual describe en términos operacionales cada uno de los indicadores. Decidimos manejar la versión en extenso de forma separada para facilitar la evaluación por las educadoras, como lo hacen otros instrumentos con base en rúbricas, por ejemplo, el Classroom Assessment Scoring System (Pianta, La Paro & Hamre, 2008).

Tabla 6. Ejemplo de rúbrica en extenso del ítem de conversación

Descriptorios de indicadores por nivel	Indicadores agrupados		
	Concretar y congruencia	Gramática, semántica y organización de información	Aportación de información
Nivel 2	Divaga en los comentarios que realiza Se limita a afirmar, negar o ceder a lo que dijeron otros	La estructura (gramática) es incorrecta Se le dificulta aclarar u organizar ideas o información de forma congruente, aunque se le apoye	Le cuesta aportar información nueva o prefiere hablar de un mismo tema
Nivel 4	Los comentarios que realiza son congruentes, pero se le dificulta concretar sus aportaciones	Ha mejorado la estructura de los comentarios que realiza Organiza ideas o información, pero de forma limitada o con errores	Conversa sobre experiencias similares a las que acaban de manifestar sus compañeros Aporta comentarios redundantes
Nivel 6	Ajusta el lenguaje al público Las respuestas o comentarios que realiza son congruentes (por tipo de respuesta y tema), concretos y realiza precisiones de estos (como pedir que se le aclare una pregunta, reformular una pregunta) con las preguntas	Es capaz de organizar ideas o información que se le presenta, en su mayoría sin errores La estructura sintáctica y gramatical de sus aportaciones es correcta (tiempos verbales, conectores, orden, otros)	Inicia conversaciones de diferentes experiencias, sucesos o temas En la resolución de problemas toma en cuenta los comentarios de sus compañeros para proponer soluciones que se ajusten al problema Aporta nueva información o diferentes puntos de vista del mismo tema

Asimismo, en el manual se agregaron ejemplos comunes de desempeño de los niños para cada uno de los niveles clave de la rúbrica; estos tienen la intención de facilitar la ubicación del desempeño de los alumnos.

Tabla 7. Ejemplo de desempeño del aspecto de conversación

Nivel 2	Alumnos que se expresan con monosílabos Rehúyen conversar, aunque se les insista, porque aún no han consolidado la habilidad para conversar (darse a entender, comprender al otro) Se expresan mayormente con apoyo de señas Señalan algunos objetos porque no sabe cómo se llaman Se le dificulta pronunciar palabras
---------	--

Nivel 4	<p>Alumnos que ya participan por sí mismos</p> <p>Aunque pronuncian incorrectamente las palabras, pero logran darse a entender; por ejemplo, los niños que dicen “sabo”, “quiedo”, otras</p> <p>Alumnos que utilizan frases como “mañana miré una película”</p> <p>Alumnos que usan lo que podría considerarse un exceso de muletillas, por ejemplo “y fuimos y le dije y corrimos y jugamos y...”</p> <p>Alumnos que han aumentado su vocabulario, pero que se expresan sin una organización de las ideas</p>
Nivel 6	<p>Alumnos que se comunican con frases largas y elaboradas si es necesario, como “ayer fuimos al cine con mis abuelos”</p> <p>Alumnos que han incrementado su vocabulario</p>

Por último, cada aspecto de la rúbrica se acompaña de ejemplos de actividades que las educadoras pueden utilizar durante la aplicación; tienen el objetivo de modelar situaciones propicias para recolectar cierto tipo de información; por ejemplo, para conversación se proponen actividades como preguntar que hicieron ayer, observar a los alumnos mientras hablan entre ellos en actividades cotidianas, como el receso y platicar con el alumno cuando este se acerca por propia iniciativa. Por último, se agrega un glosario de términos que pudieran parecer ambiguos; su objetivo es evitar confusiones y facilitar la asimilación del instrumento.

Análisis descriptivos de los resultados obtenidos a través de la aplicación de la rúbrica

A partir de los resultados de la tabla 8, advertimos que hay un dominio general; si bien los desempeños observados abarcan todos los niveles, para el lenguaje oral, los niveles cinco y seis concentraron la mayor cantidad de casos. Por su parte, el ámbito del lenguaje escrito tuvo los porcentajes mayores en el nivel dos, tanto en lo relativo a la producción como a la interpretación. Lo anterior se traduce como un dominio general mayor de los evaluados en la oralidad, mientras que se evidencia uno menor en los aspectos referentes al dominio de la lengua escrita.

Aunado a lo anterior, los resultados por grado escolar reflejan un nivel de desempeño de los alumnos evaluados consecuente con el grado escolar al que pertenecían; es decir, los alumnos de tercero de preescolar tendieron a ubicarse por encima del desempeño de los alumnos de segundo grado.

Tabla 8. Distribución del nivel de dominio total y por grado escolar

Ítem	Gra- do	Nivel							Total
		1	2	3	4	5	6	7	
Conver- sación	To- tal	2.1%	5.4%	14.1%	21.2%	18.3%	29.5%	9.5%	100%
	2do.	3.6%	10.8%	17.1%	28.8%	17.1%	22.5%	0.0%	100%
	3ro.	0.8%	0.8%	11.5%	14.6%	19.2%	35.4%	17.7%	100%

Narración	Total	4.1%	14.9%	17.8%	18.7%	18.3%	<i>24.1%</i>	2.1%	100%
	2do.	9.0%	27.9%	32.4%	19.8%	8.1%	2.7%	0.0%	100%
	3ro.	0.0%	3.8%	5.4%	17.7%	26.9%	<i>42.3%</i>	3.8%	100%
Descripción	Total	6.2%	3.7%	11.6%	20.7%	22.4%	<i>34.9%</i>	0.4%	100%
	2do.	12.6%	7.2%	18.9%	28.8%	19.8%	12.6%	0.0%	100%
	3ro.	0.8%	0.8%	5.4%	13.8%	24.6%	<i>53.8%</i>	0.8%	100%
Explicación	Total	3.7%	9.5%	18.3%	22.8%	20.7%	<i>24.5%</i>	0.4%	100%
	2do.	7.2%	14.4%	28.8%	23.4%	18.9%	7.2%	0.0%	100%
	3ro.	0.8%	5.4%	9.2%	22.3%	22.3%	<i>39.2%</i>	0.8%	100%
PLE	Total	26.1%	31.1%	11.2%	15.4%	9.1%	5.8%	1.2%	100%
	2do.	49.5%	41.4%	0.0%	9.0%	0.0%	0.0%	0.0%	100%
	3ro.	6.2%	22.3%	20.8%	20.8%	16.9%	10.8%	2.3%	100%
ILO	Total	2.1%	9.1%	17.0%	22.4%	27.8%	21.2%	0.4%	100%
	2do.	4.5%	18.0%	27.9%	23.4%	19.8%	6.3%	0.0%	100%
	3ro.	0.0%	1.5%	7.7%	21.5%	34.6%	33.8%	0.8%	100%
ILE	Total	12.0%	27.4%	17.0%	17.8%	12.0%	13.3%	0.4%	100.0%
	2do.	20.7%	45.9%	22.5%	10.8%	0.0%	0.0%	0.0%	100%
	3ro.	4.6%	11.5%	12.3%	23.8%	22.3%	<i>24.6%</i>	0.8%	100%

ILO: interpretación de la lengua oral
 PLE: producción de la lengua escrita
 ILE: interpretación de la lengua escrita

Nota: Las cifras resaltadas en cursivas corresponden a los porcentajes más altos.

Evidencias de validez

Basadas en el contenido. A continuación, presentamos los resultados obtenidos a través del juicio de expertos.

Comité de jueceo de diseño. Los jueces aprobaron las características generales del instrumento, hicieron sugerencias a la estructura propuesta y señalaron qué podría adecuarse a lo estipulado en la propuesta curricular (SEP, 2017, p.180), en relación con la noción de práctica social del lenguaje y sus tres ejes articuladores: producción, interpretación y reflexión sobre la lengua, a su vez, divididas en los ámbitos lengua oral y lengua escrita.

Posteriormente, durante el trabajo con el comité de indicadores, se sugirió eliminar la dimensión de reflexión sobre la lengua, debido a que esta se encuentra implícita dentro de las dimensiones de producción e interpretación sobre la lengua (Ferreiro, 2006). Con base en estas recomendaciones, obtuvimos la estructura final que mostramos en la tabla 4.

Comité de jueceo del instrumento. Los jueces valoraron de forma adecuada el contenido de los ítems, pues el coeficiente de validez de contenido (CVC) para cada aspecto evaluado bajo los criterios agrupados de calidad, relevancia, congruencia y suficiencia resultó con valores cercanos o superiores a .90.

Tabla 9. Coeficiente de acuerdo entre jueces y de validez de contenido

Aspecto evaluado	CVct
Conversación	0.91
Narración	0.91
Descripción	0.83
Explicación	0.87
Interpretación de la lengua oral	0.95
Producción de la lengua escrita	0.95
Interpretación de la lengua escrita	0.87

Nota: Valores entre .8 y .9 = buenos, por encima de .9 = excelentes (Hernández-Nieto, 2011).

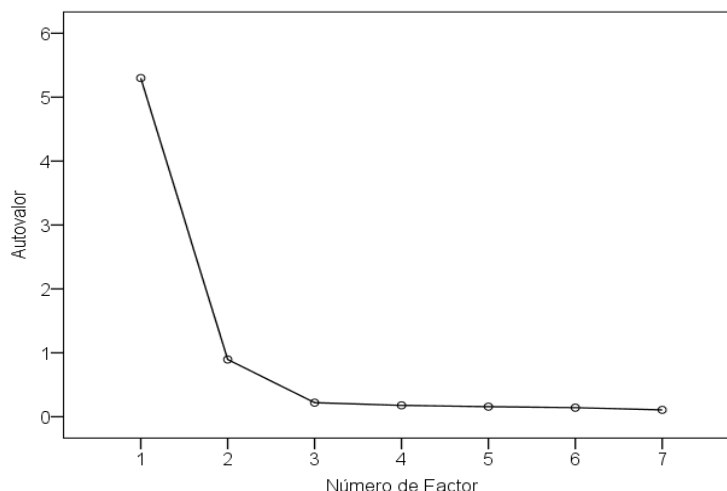
Aunado a lo anterior, la valoración cualitativa de los jueces se orientó mayormente a mejorar la claridad en la redacción de algunos niveles de la rúbrica y a clarificar la gradación de ciertos niveles de desempeño. Los apartados referentes a descripción y narración no recibieron comentarios. Por último, las aportaciones de este comité se reflejaron en el aumento de cinco a siete niveles de desempeño, así como en la mejora de la claridad en la redacción de los niveles señalados por los expertos.

Basadas en la estructura interna

Como primer paso, comprobamos que se cumplen los supuestos para el AFE al obtener un índice de Kaiser-Meyer-Olkin de .898, así como valores significativos en la prueba de esfericidad de Bartlett, lo que representa que las varianzas de los factores se encuentran correlacionadas.

La aplicación del AFE permitió determinar la unidimensionalidad del constructo medido, cuyo único factor explica el 76% de la varianza de los ítems (ver gráfico). Con estos datos se demuestra que todos los ítems forman parte de un mismo constructo latente sobre lenguaje y comunicación.

Gráfico de sedimentación



Asimismo, todas las cargas de los ítems presentaron un nivel de saturación suficientemente alto, por lo que ninguno debió ser eliminado o sometido a revisión (ver tabla 10). Por último, obtuvimos un coeficiente de alfa de Cronbach de 0.944. Lo anterior se traduce como una confiabilidad alta de los puntajes del instrumento dado que el coeficiente se acerca a uno (Morales, 2013).

Tabla 10. Matriz factorial

	Factor
	1
Descripción	.925
Narración	.921
Explicación	.915
Interpretación de la lengua oral	.911
Conversación	.843
Interpretación de la lengua escrita	.694
Producción de la lengua escrita	.678

Nota: AFE con método de extracción: máxima verosimilitud.

DISCUSIÓN Y CONCLUSIONES

El propósito de esta investigación fue diseñar, desarrollar y obtener evidencias de validez de un instrumento de evaluación de lenguaje para alumnos de preescolar. A continuación, describimos los atributos del instrumento diseñado.

Diseño y desarrollo del instrumento

Para la construcción de la rúbrica, hicimos una adaptación de los componentes para el desarrollo de instrumentos de evaluación propuestos por Lane y colaboradores (2016), que, a su vez, se alinean con los estándares de la AERA/APA/NCME (2014). Dado lo anterior, la rúbrica construida se encuentra formulada con base en directrices reconocidas en el campo del desarrollo de instrumentos.

La rúbrica diseñada integra atributos que la vuelven atractiva para su uso en el sistema educativo mexicano. Estos atributos son: está armonizado con el currículo nacional; representa los distintos contenidos del currículo tanto de la lengua oral como escrita; permite evaluar desempeños complejos; y tiene una forma de administración fácil y confiable.

La rúbrica está adaptado al currículo nacional y se diseñó tomando como referente el PEP 2017. Ello ofrece la ventaja de utilizar el mismo lenguaje pedagógico que las educadoras, lo que significa, por una parte, que no es un contenido disruptivo y, por otra, se capitaliza que las educadoras están formadas en el currículo y orientan su práctica docente a partir de este. La enseñanza es una acción sistémica, dependiente del contexto, por lo que la consideración de este en las evaluaciones condiciona necesariamente su validez (Jornet, Perales y González, 2020).

La rúbrica permite evaluar de forma equilibrada los distintos contenidos del currículo, por ejemplo, se da más prioridad a la oralidad, la cual está representada con mayor relevancia en el currículo (SEP, 2017). Por una parte, la rúbrica vincula cada uno de los ítems a alguno de los aprendizajes esperados del programa; asimismo, desglosa los aprendizajes esperados en indicadores identificables y gradados por nivel de desempeño. De esta manera, se evita el error de concentrarse en la evaluación de aspectos de la lengua escrita: lectura y escritura.

El instrumento, al ser una rúbrica, posibilita evaluar desempeños como las competencias que espera desarrollar el PEP 2017. Para el instrumento, diseñamos una rúbrica del tipo analítica y general (Arter, 2010), es decir, que es aplicable en diversos desempeños de los niños, por lo que no se requiere que las educadoras implementen tareas evaluativas específicas. Además, las rúbricas son un instrumento capaz de promover el aprendizaje de niños y niñas, pues posibilitan dos acciones necesarias para la evaluación formativa: reconocer el desempeño actual de los alumnos y orientar respecto a desempeños subsiguientes sobre el que ellos pueden avanzar (Andrade, Bennett & Cizek, 2019; Arter, 2010; Black & William, 2018; Hattie & Timperley, 2007; Martínez, 2012a, 2012b).

Consideramos una fortaleza que la evaluación se realice por la educadora, ya que su incorporación ayuda a incrementar la precisión. Ella es quien conoce el rendimiento general del alumno a partir de que hace un monitoreo y seguimiento de las actividades de los niños. Por una parte, como vimos en la introducción de este artículo, se evita el sesgo de un evaluador externo (Leyva, 2011; Moreno, 2015). La aplicación de instrumentos basados en rúbricas por docentes de preescolar ha sido utilizada por otros instrumentos de evaluación (López & Willms, 2020; Marín, Guzmán y Castro, 2012; Willms, 2018) y su incorporación es recomendada como uno de los principios de evaluación para los menores de seis años (Shepard et al., 1998). Algunas investigaciones han encontrado que las evaluaciones de los docentes mediante instrumentos basados en rúbricas tienen valores similares a los aportados por evaluadores externos (Willms, 2014), lo cual indica que las evaluaciones con base en rúbricas administradas por docentes pueden tener resultados relativamente confiables.

De igual modo, el instrumento incorpora características que contribuyen a la fácil asimilación y aplicación por el propio docente. Su diseño tiene una cantidad manejable de ítems, y la facilidad de poder aplicarlo en distintas situaciones pedagógicas,

asignaturas y de forma gradual. Además, su versión impresa y en línea pueden utilizarse en conjunto o por separado; la ventaja de la aplicación en línea estriba en que es capaz de resguardar los datos y proveer informes de los resultados en formato individual. A futuro, se espera generar informes por grupo, escuela, zona o municipio. En este sentido, el docente tiene acceso a los informes de resultados inmediatamente al término de la evaluación, lo que, de acuerdo con el INEE (s.f.), contribuye a incrementar la posibilidad de que docente y alumno se beneficien de sus propios resultados.

Evidencias de validez basadas en el contenido

De forma general, para sustentar el instrumento, creamos elementos que son considerados por la AERA/APA/NCME (2014) como fuentes de evidencia de validez de contenido: estructura de la rúbrica, operacionalización de las variables, especificaciones de contenido y jueceo de la operacionalización de las variables y de la rúbrica. Respecto al jueceo de la rúbrica, se obtuvieron coeficientes calificados de buenos a excelentes para cada uno de los ítems (Hernandez-Nieto, 2011).

Una de las estrategias principales de esta investigación consistió en focalizar los esfuerzos en obtener evidencias de validez del contenido, ya que dentro de la literatura estas se reconocen como un requisito previo a la adquisición de evidencias de cualquier otro tipo (AERA/APA/NCME, 2014; Vogt, King & King, 2004). Sin embargo, este requisito no siempre se cumple, y un ejemplo son los instrumentos publicados de 2000 a 2010 en la revista *Journal of Educational Psychology*, pues solo el 16% de los artículos concernientes a la construcción de instrumentos reportaron evidencias referentes a aspectos de contenido (Collie & Zumbo, 2014).

Aunado a lo anterior, acorde con los señalamientos de Brijmohan et al. (2018), dedicamos esfuerzos para asegurar que los expertos participantes en el proceso de diseño y de valoración de la rúbrica contaran con la pericia necesaria en evaluación, diseño de instrumentos y en el contenido. Asimismo, buscamos que tuvieran discernimiento de los conocimientos y las habilidades propuestos en el currículo mexicano, en la teoría subyacente y experiencia práctica respecto a los aprendizajes y dificultades de los estudiantes, así como la relación de estos aspectos con sus conocimientos y habilidades. Según Jornet, González y Suárez (2010), el dedicar especial atención a la conformación de los comités dota de credibilidad y utilidad al instrumento desarrollado.

Evidencias de validez basadas en la estructura interna

Esta fuente de evidencia se obtuvo a través del análisis de confiabilidad de alfa de Cronbach y del AFE, que arrojaron un coeficiente de confiabilidad de 0.944 y un factor único que explica más del 76% de la varianza de los puntajes alcanzados. En el campo de las ciencias sociales, ambas cifras son consideradas como altas (Morales, 2013).

En este trabajo presentamos dos fuentes de validez que respaldan el instrumento desarrollado para los usos pretendidos. Cabe mencionar que solo el 21% de los instrumentos publicados en la revista *Journal of Educational Psychology* entre 2000 y 2010 reportaron esta cantidad de fuentes (Collie & Zumbo, 2014). Por su parte, el 62% de los instrumentos de evaluación del lenguaje hacen explícitos datos sobre

confiabilidad; sin embargo, no todos los artículos atendieron apropiadamente los criterios establecidos en la literatura. Asimismo, apenas el 54% mostraron información referente a aspectos de confiabilidad y validez (Castro y Neira, 2008).

Limitaciones

Para dar mayor soporte a los usos pretendidos del instrumento: formativo y de seguimiento de tendencias, será necesario que se generen dos tipos adicionales de evidencias de validez: del proceso de respuesta y de las consecuencias. Sobre el primero, es deseable que se determine el grado de asociación que tiene el instrumento cuando este es aplicado por un evaluador externo y la educadora frente a grupo. Ello implicaría tener observadores en aula para identificar los desempeños de los alumnos. Además, se requiere identificar el empleo que las educadoras hacen de los resultados, en específico si se utilizan para un uso formativo: de retroalimentación a los alumnos y de modificación de las prácticas docentes, así como si existen consecuencias no deseadas a partir del uso del instrumento.

Es necesario que el desarrollo de instrumentos de evaluación incorpore la perspectiva cualitativa en su diseño. La construcción de este instrumento privilegió el análisis documental y el trabajo con jueces expertos. Se pudo enriquecer con otras perspectivas cualitativas, por ejemplo, de la observación en el aula y la perspectiva de las educadoras en su experiencia de uso del instrumento. Con base en las necesidades emergentes durante este estudio, reconocemos la pertinencia de metodologías mixtas (Creswell & Creswell, 2018) para el desarrollo de instrumentos de evaluación, en especial para aquellos con una lógica similar al aquí propuesto.

REFERENCIAS BIBLIOGRÁFICAS

- American Educational Research Association/American Psychological Association/ National Council on Measurement in Education (AERA/APA/NCME) (2014). *Standards for educational and psychological testing*. Estados Unidos.
- American Psychological Association-Coalition for Psychology in Schools and Education (2019). *Top 20 principles from psychology for early childhood teaching and learning*. <https://www.apa.org/ed/schools/teaching-learning/top-twenty-early-childhood.pdf>
- Andrade, H., Bennett, R. & Cizek, G. (2019). Prefacio. En H. L. Andrade, R. Bennett y G. Cizek. *Handbook of formative assessment in the disciplines*. Londres: Routledge.
- Arter, J. (2010). Scoring rubrics. En P. Peterson, E. Baker y B. McGaw. *International Encyclopedia of Education* (3a ed.) (pp. 123-139). Reino Unido: Elsevier.
- Bentler, P. (2006). *EQS 6 Structural Equations Programs Manual*. California, EUA: Multivariate Software Inc.
- Black, P. & Wiliam, D. (2019). Developing a theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, vol. 21, núm. 1, pp. 5-31. <https://doi.org/10.1007/s11092-008-9068-5>
- Black, P. & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, vol. 25, núm. 6, pp. 551-575. <https://doi.org/10.1080/0969594X.2018.1441807>

- Brijmohan, A., Khan, G. A., Orpwood, G., Sandford Brown, E. & Childs, R. A. (2018). Collaboration between content experts and assessment specialists: Using a validity argument framework to develop a college mathematics assessment. *Canadian Journal of Education*, vol. 4, núm. 2, pp. 584-600. <https://journals.sfu.ca/cje/index.php/cje-rce/article/view/3239>
- Cardemil, C. y Román, M. (2014). Presentación de la sección temática: la importancia de analizar la calidad de la educación en el nivel inicial y preescolar. *Revista Iberoamericana de Evaluación Educativa*, vol. 7, núm. 1, pp. 9-11. <https://revistas.uam.es/riee/issue/view/423/254>
- Castro, F. y Neira, L. I. (2008). Caracterización de instrumentos de evaluación del desarrollo del lenguaje para hablantes del español. *Revista Areté*, vol. 8, núm. 1, pp. 53-62. <https://arete.iberu.edu.co/article/view/435>
- Collie, R. & Zumbo, B. (2014). Validity evidence in the journal of educational psychology: Documenting current practice and a comparison with earlier practice. En R. Collie & E. K. H. Chan (eds.). *Validity and validation in social, behavioral, and health sciences* (pp. 113-135). Nueva York: Springer International Publishing.
- Creswell, J. & Creswell, D. (2018). *Research design: Qualitative, quantitative, and mixed methods approach* (5a ed.). Estados Unidos: SAGE Publishing.
- Dubeck, M. & Gove, A. (2015). The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations. *International Journal of Educational Development*, vol. 40, pp. 315-322. <https://doi.org/10.1016/j.ijedu-dev.2014.11.004>
- Fernald, L., Prado, E., Karinger, P. & Raikes, A. (2017). *A toolkit for measuring early childhood development in low- and middle-income countries*. Washington, DC: The World Bank.
- Ferreiro, E. (2006). La escritura antes de la letra. *CPU-e, Revista de Investigación Educativa*, núm. 3, pp. 1-52. <https://doi.org/10.25009/cpue.v0i3.136>
- García Medina, A. M., Aguilera García, M. A., Pérez Martínez, M. G. y Muñoz Abundez, G. (2011). *Evaluación de los aprendizajes en el aula. Opiniones y prácticas de docentes de primaria en México*. México: INEE.
- Gómez Meléndez, L. E., Cáceres Mesa, M. L. y Zúñiga Rodríguez, M. (2018). La evaluación del aprendizaje en la educación preescolar. Aproximación al estado del conocimiento. *Revista Conrado*, vol. 14, núm. 62, pp. 247-255. <https://conrado.ucf.edu.cu/index.php/conrado/article/view/712>
- Gómez Patiño, R. y Seda Santana, I. (2008). Creencias de las educadoras acerca de la evaluación de sus alumnos preescolares: un estudio de caso. *Perfiles Educativos*, vol. 30, núm. 119, pp. 33-54.
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, vol. 77, núm. 1, pp. 81-112. <https://doi.org/10.3102/003465430298487>
- Hernández-Nieto, R. (2011). *Instrumentos de recolección de datos en ciencias sociales y ciencias biomédicas*. Mérida, Venezuela: Universidad de Los Andes.
- Hernández-Nieto, R. (2002). *Contribuciones al análisis estadístico*. Mérida, Venezuela: Universidad de Los Andes.
- Instituto Nacional de Evaluación Educativa (INEE) (2014). *El aprendizaje en preescolar en México. Informe de resultados EXCALE 00 aplicación 2011 lenguaje y comunicación y pensamiento matemático*. México.

- Instituto Nacional de Evaluación Educativa (INEE) (s.f.). *Cuadernillo técnico de evaluación educativa. Uso de resultados y retroalimentación*. México.
- Jornet Meliá, J. M., González Such, J. y Suárez Rodríguez, J. (2010). Validación de los procesos de determinación de estándares de interpretación (EE) para pruebas de rendimiento educativo. *Estudios sobre Educación*, vol. 19, pp. 11-29. <https://hdl.handle.net/10171/18323>
- Jornet Meliá, J. M., Perales Montolío, M. J. y González Such, J. (2020). El concepto de validez de los procesos de evaluación de la docencia. *Revista Española de Pedagogía*, vol. 78, núm. 276, pp. 233-252. <https://doi.org/10.22550/REP78-2-2020-01>
- Lane, S., Raymond, M., Haladyna, T. y Downing, S. (2016). Test development process. En S. Lane, M. Raymond y T. Haladyna (eds.). *Handbook of test development* (vol. II, pp. 3-18). Nueva York, EUA: Routledge.
- Leyva Barajas, Y. E. (2011). Una reseña sobre la validez de constructo de pruebas referidas a criterio. *Perfiles Educativos*, vol. 33, núm. 131, pp. 131-154. <https://doi.org/10.22201/iissue.24486167e.2011.131.24238>
- Lloret Segura, S., Ferreres Traver, A., Hernández Baeza, A. y Tomás Marco, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Anales de Psicología*, vol. 30, núm. 3, pp. 1151-1169. <https://dx.doi.org/10.6018/analesps.30.3.199361>
- López García, A. Y. & Willms, J. D. (2020). A national evaluation of kindergarten outcomes: Findings from Uruguay. En J. Hall, A. Lindorff & P. Greet (eds.). *International Perspectives in Educational Effectiveness Research* (pp. 361-381). Nueva York, EUA: Springer International Publishing.
- Marín Uribe, R., Guzmán Ibarra, I. y Castro Aguirre Guzmán, G. (2012). Diseño y validación de un instrumento para la evaluación de competencias en preescolar. *Revista Electrónica de Investigación Educativa*, vol. 14, núm. 1, pp. 182-202. <https://redie.uabc.mx/redie/article/view/308>
- Martínez Rizo, F. (2012a). Dificultades para implementar la evaluación formativa. Revisión de literatura. *Perfiles Educativos*, vol. 35, núm. 139. <https://doi.org/10.22201/iissue.24486167e.2013.139.35716>
- Martínez Rizo, F. (2012b). *La evaluación en el aula. Promesas y desafíos de la evaluación formativa*. Aguascalientes: Universidad Autónoma de Aguascalientes.
- Morales Vallejo, P. (2013). El análisis factorial en la construcción e interpretación de tests, escalas y cuestionarios. *Revista Comillas*, pp. 1-46.
- Moreno Olivo, T. (2015). Las competencias del evaluador educativo. *Revista de la Educación Superior*, vol. 44, núm. 174, pp. 101-126. <https://doi.org/10.1016/j.resu.2015.04.011>
- Pedroza Zúñiga, L. H. y Luna Serrano, E. (2017). Desarrollo y validación de un instrumento para evaluar la práctica docente en educación preescolar. *Revista Iberoamericana de Evaluación Educativa*, vol. 10, núm. 1. <https://doi.org/10.15366/riee2017.10.1.006>
- Pianta, R., La Paro, K. & Hamre, B. K. (2008). *Classroom assessment scoring system: Manual K-3*. Estados Unidos: Paul H Brookes Publishing.
- Secretaría de Educación Pública (SEP) (2018). *Manual de exploración de habilidades básicas en lectura, escritura y conteo. Herramienta para la escuela*. México: Dirección General de Desarrollo de la Gestión e Innovación Educativa.

- Secretaría de Educación Pública (SEP) (2017). *Aprendizajes clave para la educación integral. Educación preescolar. Plan y programas de estudio, orientaciones didácticas y sugerencias de evaluación*. México.
- Shepard, Lorrie, K., Sharon, L. & Wurtz, E. (1998). *Principles and recommendations for early childhood assessments*. Washington, DC: National Education Goals Panel.
- Sireci, S. & Faulkner Bond, M. (2014). Validity evidence based on test content. *Psicothema*, vol. 26, núm. 1, pp. 100-107. <https://doi.org/10.7334/psicothema2013.256>
- Snow, C. & Van-Hemel, S. (eds.) (2008). *Early childhood assessment: Why, what, and how*. Washington, DC: The National Academies Press.
- The Learning Bar (2021). *Evaluación infantil temprana*. <https://www.earlyyearsevaluation.com/index.php/es/>
- Vogt, D., King, A. y King, L. (2004). Focus groups in psychological assessment: Enhancing content validity by consulting members of the target population. *Psychological Assessment*, vol. 16, núm. 3, pp. 231-243. <https://doi.org/10.1037/1040-3590.16.3.231>
- Willms, J. D. (2018). *Las brechas de aprendizaje: uso de datos para formular la política educativa*. Francia: Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura.
- Willms, J. D. (2014). *Informe final de resultados de la evaluación infantil temprana. Estudio piloto en Colonia, Uruguay*. Fredericton: The Learning Bar Inc.
- Wortham, S. & Hardin, B. (2016). *Assessment in early childhood education* (8a ed.). Estados Unidos: Pearson.