



Magis. Revista Internacional de Investigación en Educación

ISSN: 2027-1174

ISSN: 2027-1182

articulosmagis@gmail.com

Pontificia Universidad Javeriana

Colombia

Orellana-García, Pelusa; Valenzuela-Hasenohr, María
Francisca; Kung, Melody; Elmore, Jeff; Stenner, A. Jackson
EVOC: un Instrumento para Evaluar Vocabulario en Español en Chile
Magis. Revista Internacional de Investigación en Educación, vol. 13, 2020, Julio-Junio, pp. 1-26
Pontificia Universidad Javeriana
Bogotá, Colombia

DOI: <https://doi.org/10.11144/Javeriana.m13.eiev>

Disponible en: <https://www.redalyc.org/articulo.oa?id=281068067014>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en [redalyc.org](https://www.redalyc.org)

[redalyc.org](https://www.redalyc.org)

Sistema de Información Científica Redalyc

Red de Revistas Científicas de América Latina y el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso
abierto

EVOC: un Instrumento para Evaluar Vocabulario en Español en Chile

EVOC: An Instrument to Evaluate the Spanish Lexicon in Chile

Artículo de investigación | Research Article

Fecha de recepción: 5 de agosto de 2018

Fecha de aceptación: 22 de marzo de 2019

Fecha de disponibilidad en línea: marzo de 2020

doi: 10.11144/Javeriana.m13.eiev

PELUSA ORELLANA-GARCÍA

porellan@uandes.cl

UNIVERSIDAD DE LOS ANDES, CHILE

<https://orcid.org/0000-0003-3851-8780>



JEFF ELMORE

jelmore@lexile.com

METAMETRICS, INC., ESTADOS UNIDOS

<https://orcid.org/0000-0002-7823-2704>



MARÍA FRANCISCA VALENZUELA-HASENOHR

mvalenzuela@uandes.cl

UNIVERSIDAD DE LOS ANDES, CHILE

<https://orcid.org/0000-0002-7019-9110>



A. JACKSON STENNER

jstenner@lexile.com

THE UNIVERSITY OF NORTH CAROLINA

AT CHAPEL HILL, ESTADOS UNIDOS

<https://orcid.org/0000-0002-6648-9895>

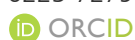


MELODY KUNG

mkung@purdue.edu

PURDUE UNIVERSITY, INDIANA

<https://orcid.org/0000-0002-6223-7279>



Para citar este artículo | To cite this article

Orellana-García, P.; Valenzuela-Hasenohr, M. F.; Kung, M.; Elmore, J. & Stenner, A. J. (2020). EVOC: un Instrumento para Evaluar Vocabulario en Español en Chile. *magis, Revista Internacional de Investigación en Educación*, 13, 1-26. doi: 10.11144/Javeriana.m13.eiev



Resumen

Este estudio describe la creación y validación de una prueba adaptativa y autoadministrada de vocabulario en español para estudiantes de kínder a octavo año básico. Usando un *corpus* de un millón (1.000.000) de palabras, provenientes de textos escolares, se seleccionaron palabras de distinta frecuencia y dificultad. Se confeccionaron ítems siguiendo el formato del test PPVT [Test de Vocabulario en Imágenes Peabody, Peabody Picture Vocabulary Test] que representa cada palabra mediante fotografías. Los distractores se originan aleatoriamente desde el mismo *corpus*. Se estableció la validez del constructo, concurrente y con consistencia interna. EVOC es un instrumento válido y confiable para evaluar vocabulario en español que permite adaptar las prácticas pedagógicas a las necesidades de los alumnos.

Palabras clave

Vocabulario; comprensión; evaluación; semántica

Abstract

This study describes the creation and validation of an adaptive and self-applied test of lexicon in Spanish for kindergarten to eighth-grade basic school students. Using a corpus of 1,000,000 words from school texts, some words with different frequency and difficulty were selected. The test items were prepared according to the test format PPVT [Peabody Picture Vocabulary Test], representing each word with pictures. The distractors were generated randomly from the same corpus. The construct validity, concurrent validity and internal consistency were determined. EVOC is a valid and reliable instrument to evaluate the lexicon in Spanish that allows adapting the pedagogical practices to the student's needs.

Keywords

Vocabularies; comprehension; evaluation; semantics

Descripción del artículo | Article description

Este artículo de investigación, derivado del proyecto *Validación y empaquetamiento de la plataforma Dialect*, da cuenta de los procesos de diseño y validación de la prueba EVOC, un componente de la plataforma Dialect para diagnóstico de habilidades tempranas.

Introducción

El uso de evaluaciones estandarizadas de vocabulario permite entregar información valiosa sobre los conocimientos léxicos de los estudiantes. Esta ayuda a los docentes a planificar las actividades pedagógicas en forma adecuada (Dickinson & Tabors, 2005; Izura, Cuetos & Brysbaert, 2014). Por lo general, las evaluaciones de vocabulario contienen normas, para comparar el nivel de vocabulario de un estudiante con el de sus pares del mismo grupo etario. Esta información permite desarrollar intervenciones pedagógicas individualizadas para aquellos estudiantes que aún no alcanzan las metas específicas, o que tienen un nivel de vocabulario insuficiente para enfrentar tareas de comprensión o de escritura.

El objetivo de este estudio fue desarrollar una prueba en formato digital y adaptativa que midiera el nivel de vocabulario receptivo en español para estudiantes de kínder a octavo año básico. Pese a que hay algunos instrumentos de evaluación del vocabulario, ellos se basan en *corpus* de diccionarios de español en general (e.g., TEVI-R, Echeverría, Herrera & Segure, 2002) y no en las palabras que provienen de los textos a los cuales se exponen los alumnos durante la enseñanza preescolar y la enseñanza básica. El vocabulario contenido en las lecturas que los estudiantes realizan durante su vida escolar constituye una base fundamental para comprender, procesar, interpretar y elaborar textos como adultos. Por otra parte, en el contexto hispanohablante hay escasez de evaluaciones de vocabulario en formato digital, que faciliten la aplicación masiva de la medición y permitan obtener resultados de manera oportuna y costo eficiente (Neira & Castro, 2008; Pearson, Hiebert & Kamil, 2007). Instrumentos como el Test de Vocabulario en Imágenes de Peabody (Dunn, 1986), o el Woodcock-Muñoz (Muñoz-Sandoval, Woodcock, McGrew & Mather, 2005) son usados en Chile, pero algunos de ellos requieren capacitación y certificación especial, lo que los hace costosos. A ello se agrega que, al necesitar de corrección individual en lápiz y papel, la información que se obtiene a partir de su aplicación es lenta y poco eficiente, ya que además se administra en forma individual.

Por otra parte, los avances tecnológicos se han ido aprovechando para ser usados como contextos en los cuales evaluar las habilidades lingüísticas de los estudiantes (e.g., Schaefer, Bowyer-Crane, Herrmann & Fricke, 2016). El uso de instrumentos de evaluación en soporte tecnológico permite la adaptación a variantes regionales, según el dialecto que se utilice, para representar la terminología local de manera más precisa. El instrumento que se propone cumple con esta característica.

Por otra parte, contar con instrumentos que permitan determinar el vocabulario de los alumnos es primordial en América Latina, donde los alumnos con un vocabulario descendido muchas veces no son identificados, y donde el desarrollo de la comprensión lectora es insuficiente para las actuales demandas de procesamiento de texto escrito (Fien, Santoro, Baker, Park, Chard, Williams & Haria, 2011).

¿Qué se entiende por vocabulario?

Conocer una palabra es entender sus significados (Schatschneider, Harrell & Buck, 2007). Richard K. Wagner, Andrea E. Muse y Kendra R. Tannenbaum (2007) distinguen tres dimensiones: *amplitud*, que se refiere al tamaño del léxico mental, es decir, la cantidad de palabras que maneja un individuo; *profundidad*, que se refiere a la riqueza del significado; y la *fluidez de acceso*, que se define como la velocidad con la cual el individuo accede al significado de la palabra. Este conocimiento considera si la palabra se usa en contextos formales o más informales, las formas gramaticales y connotaciones afectivas (Nagy & Scott, 2000). El vocabulario incluye aquellas palabras que debemos saber para comunicarnos de manera efectiva, ya sea al hablar o escribir (vocabulario expresivo) como al escuchar o leer (vocabulario receptivo) (Neuman & Dwyer, 2009).

Las palabras permiten conectar a la persona con experiencias previas y su uso depende de los contextos en que estén (Pearson, Hiebert & Kamil, 2007). William E. Nagy y Judith A. Scott (2000) identificaron cinco aspectos del conocimiento de una palabra necesarios para la lectura: (a) *incrementalidad*, es decir, conocer una palabra no es algo de todo o nada, sino que cada encuentro del sujeto con la palabra permite avanzar en la precisión y flexibilidad de su uso; (b) *multidimensionalidad*, que incluye conocimiento cualitativamente distinto para tener un conocimiento más profundo; (c) *polisemia*, pues muchas palabras tienen diversos significados; (d) *interrelación*, que evidencia que aprender una palabra implica realizar asociaciones con otras, en un contexto lingüístico determinado; y (e) *heterogeneidad*, que indica que el significado de una palabra cambia según su

función y estructura. Su enseñanza directa debe ir más allá de memorizar una definición e incluye la exposición en forma repetida a ella y la entrega de información contextual (John Osborn & Bonnie B. Armbruster, 2001, en Phythian-Sence & Wagner, 2007).

La importancia del vocabulario en la habilidad lectora

La contribución del conocimiento léxico en la comprensión lectora está ampliamente documentada en la literatura (NICHD-NIH, 2000; Shany & Biemiller, 2010; Verhoeven & Van Leeuwe, 2008). Se ha demostrado la existencia de una relación bidireccional entre vocabulario y comprensión (Beck, Perfetti & McKeown, 1982), que se fortalece a lo largo del tiempo (Baumann & Kame'enui, 2004). Asimismo, la literatura ha consignado que el vocabulario es uno de los principales predictores de la comprensión lectora (Beck & McKeown, 1991; Cain & Oakhill, 2011; Quinn, Wagner, Petscher & López, 2015; Tannenbaum, Torgesen & Wagner, 2006): la *cantidad* de palabras que una persona maneja es un excelente predictor de su comprensión lectora (Protopapas, Sideridis, Mouzaki & Simos, 2007).

Además de la cantidad, profundidad y fluidez de acceso, el nivel de dificultad de una palabra es un factor que determina la complejidad de un texto (Graesser, McNamara, Cai, Conley, Li & Pennebaker, 2014; Williamson, Fitzgerald & Stenner, 2014). Aquellas lecturas donde hay una mayor proporción de palabras desconocidas para los estudiantes resultarán más difícil de comprender que aquellas donde el vocabulario es más familiar. Tradicionalmente, la frecuencia en la ocurrencia de una palabra ha sido considerada como un *proxy* para su nivel de conocimiento (e.g., Chen & Meurers, 2016; Ryder & Slater, 1988). Ello implica que aquellas palabras que aparecen más frecuentemente en un conjunto de textos serían más fáciles de comprender que otras cuyo uso es menos frecuente. Esta definición de términos es comúnmente usada en las fórmulas de lecturabilidad que buscan establecer el grado de dificultad de los textos de manera más cuantitativa.

¿Por qué evaluar el conocimiento del vocabulario?

Para poder enseñar con efectividad, es necesario plantearse previamente cómo se evalúa y cómo se desarrolla la habilidad evaluada (Pearson, Hiebert & Kamil, 2007). La evaluación del vocabulario es un proceso complejo, pues este se desarrolla en múltiples dimensiones. El estudiante debe adquirir un buen número de palabras y también los diferentes significados de las palabras que aprende. Ello incluye conocimientos de su

fonología, ortografía, morfología, uso sintáctico y pragmática, además de los propiamente semánticos. Cuando un niño aprende una nueva palabra, su conocimiento de estas distintas dimensiones no está sincronizado (Silverman & Hartranft, 2015).

La literatura señala la importancia de evaluar el vocabulario en contexto, pues su significado depende de este (Biemiller, 2006). Para el instrumento que se presenta, la imagen visual se ha considerado como el contexto que permite al sujeto asignar significado de manera más precisa; este contexto se entrega a través de una fotografía, no de un dibujo. Esta información contribuye a acercar al estudiante evaluado a la realidad y facilita su reconocimiento.

Los propósitos para evaluar vocabulario pueden ser diferentes: uno es el *tamizaje*, que permite determinar de manera gruesa los alumnos que tienen dificultades, para focalizar el apoyo pedagógico desde el inicio de la enseñanza; otro es el *diagnóstico*, que entrega información respecto de cuáles son las áreas específicas que requieren mayor apoyo; un tercer propósito es el de *monitorear el progreso*, que entrega información respecto de si los alumnos están respondiendo a la instrucción, y así ajustarla si es necesario para lograr el aprendizaje; y otro es el de *resultado final*, que indica cuánto ha crecido el vocabulario en el estudiante en un determinado período, y su relación con el aprendizaje del resto del grupo (Snow, Burns & Griffin, 1998; Walpole & McKenna, 2007).

Necesidad de acercar la evaluación a la enseñanza

Linda Darling-Hammond y John Bransford (2005) sostienen que la evaluación es una herramienta central del proceso enseñanza-aprendizaje, pues permite monitorear los logros de los alumnos y ofrecer oportunidades de aprendizaje ajustadas a las necesidades. A ello, Allison Morris (2011) agrega que el diseño y la implementación de estrategias basadas en la evaluación debe ser una práctica constante en el ámbito educativo. Un lector necesita entender el 95% de las palabras de un texto (Nagy & Scott, 2000), para comprender lo que lee; por lo tanto, si ello no se cumple, tampoco podrá aprender las palabras desconocidas a través de la lectura independiente. Por esto, conocer el nivel de vocabulario del estudiante es el punto de partida para la enseñanza del léxico en sala de clases. Una evaluación certera permite alinear los objetivos de instrucción con una pedagogía apropiada (Beck & McKeown, 2007), de acuerdo con las necesidades detectadas. Por otra parte, Cecilia Thorne, Kim Morla, Paola Uccelli, Teresa Nakano, Beatriz Mauchi, Lorena Landeo, Angie Vásquez y Román Huerta

(2013) afirman que se requiere una intervención explícita que permita expandir el vocabulario de los lectores para que puedan contar con las herramientas necesarias para comprender un texto. William E. Nagy y Judith A. Scott (2000) agregan que cuanto mayor sea el número de contextos en que una palabra pueda encontrarse, de mejor manera el alumno adquirirá su significado.

P. David Pearson, Elfrieda H. Hiebert y Michael L. Kamil (2007) sostienen que gran parte de los instrumentos que evalúan vocabulario son inadecuados para documentar la relación entre el aprendizaje de una palabra y su medida asociada a la comprensión. Ello, pues esta relación es extremadamente compleja, ya que incluye conocimiento de vocabulario, conceptual y cultural, y también oportunidades de instrucción. Los estudiantes adquieren de mejor manera el vocabulario cuando este es usado en contextos auténticos, y la imagen icónica puede considerarse uno de ellos; esta característica no incluiría cuando las palabras se presentan en formatos aislados, como las listas extensas. Los autores agregan que los distractores que se utilicen en este tipo de pruebas deben incluir: a) una palabra que tenga un significado que sea más común que la palabra clave, pero que sea descartada de acuerdo al significado según el contexto; b) una palabra que represente información correcta pero que sea incorrecta de acuerdo a la palabra clave; c) una palabra que tenga una alternativa de interpretación según el contexto de la palabra clave; y d) otras palabras que se parezcan a la palabra clave (NAGB, 2005). Por último, Pearson, Hiebert y Kamil (2007) indican que la evaluación del vocabulario comprensivo en estudiantes debe centrarse en dominios particulares de su interés, por ejemplo, las palabras que se incluyen en el currículum o en bandas de frecuencia.

Uso de la tecnología y evaluación en educación

Los avances tecnológicos han cambiado las herramientas a través de las cuales aprenden los estudiantes. También es posible extender esto a instrumentos de evaluación (Biancarosa & Griffiths, 2012). Los dispositivos digitales ofrecen una oportunidad para incorporarlos al desarrollo de la alfabetización. P. David Pearson, Elfrieda H. Hiebert y Michael L. Kamil (2007) recomiendan buscar diferentes formatos para la evaluación del vocabulario y enfatizan el uso de computadores como herramienta necesaria de incorporar en el contexto actual. Por otro lado, Ricardo Rosas, Francisco Ceric, Andrés Aparicio, Paulina Arango, Rodrigo Arroyo, Catalina Benavente, Pablo Escobar, Polín Olguín, Marcelo Pizarro, María Paz Ramírez, Marcela Tenorio y Soledad Véliz (2015) señalan que es posible evaluar

dominios cognitivos con instrumentos no tradicionales. Al presentar una tarea de manera diferente a la clásica evaluación de lápiz y papel, el estudiante no se sentiría presionado, lo cual elimina la presión y ansiedad que muchas evaluaciones tradicionales conllevan. Ello facilitaría obtener un conocimiento certero de las habilidades alcanzadas por el evaluado. El uso de evaluaciones a través de *tablets* o computadores cumple estas características, pues presentan de manera más lúdica la tarea de demostrar conocimiento del significado o uso de una palabra.

Metodología

Diseño

La construcción del instrumento contempló 7 fases. En la *primera* elegimos 24 textos escolares de Lenguaje y Comunicación, Ciencias Sociales y Ciencias Naturales de los cursos 1° a 8° año básico entregados por el Ministerio de Educación, 25 libros de cuentos de kínder y 50 textos adicionales (cuentos breves, poesías, lecturas complementarias). Ellos se digitalizaron, para generar un *corpus* de un millón de palabras. En la *segunda* fase se identificó la frecuencia de las palabras del *corpus* y se ordenaron según este criterio. De esta secuencia, se seleccionaron aleatoriamente 30 palabras por cada banda de frecuencia. Se totalizaron 330 palabras. En la *tercera* fase se escogieron 202 de las 330 palabras, según los siguientes criterios: a) que se pudieran ilustrar o graficar, b) que fueran de un nivel de dificultad medio y c) que representaran las categorías gramaticales de sustantivo, verbo, adverbio o adjetivo. En la *cuarta* fase se crearon ítems para dos formas paralelas de la prueba. Primero se ordenaron las palabras según frecuencia como un *proxy* del nivel de dificultad de la palabra y se verificó que hubiera equilibrio dentro de cada forma en cuanto a categoría gramatical. Para cada palabra clave se seleccionaron fotos en www.shutterstock.com que ilustraran cada concepto. Se usó el juicio de experto como metodología para determinar qué imagen ilustraba mejor cada concepto. Para ello, un panel de cinco académicos analizó cada placa con imágenes para asegurar fidelidad entre palabra e imagen. En la *quinta* fase se seleccionaron tres distractores para cada palabra clave, siguiendo los mismos procedimientos de la selección de las palabras clave. Posteriormente, se escogieron las imágenes para ilustrarlas desde el mismo banco de datos. Para cada ítem se construyó una placa digital en la que aparecen las cuatro ilustraciones: una para la palabra clave y una para cada uno de los tres distractores (figura 1). Las palabras se grabaron en audio, para que el alumno escuchara la palabra clave e indicara la imagen que mejor describe el concepto ilustrado en la palabra para responder el ítem.

Antes del primer piloto, ensamblamos grupos de 20 posibles distractores para 32 de los ítems de cada una de las formas. Estos distractores correspondían a palabras de las mismas bandas de frecuencia de la palabra clave de cada ítem, y a la misma categoría gramatical. Para comparar el desempeño de los estudiantes en ítems con distractores fijos y aleatorios, cada forma tuvo 32 ítems con distractores fijos y 32 con distractores aleatorios.

Figura 1

Ejemplo de ítem de la prueba EVOC. Palabra clave: lanzar



Fuente: elaboración propia con imágenes de Shutterstock

En la sexta etapa se realizó un primer piloto ($N = 571$). 392 alumnos respondieron la prueba EVOC y el TEVI-R (Echeverría, Herrera & Segure, 2002). A partir de los resultados se eliminaron los ítems que resultaron demasiado fáciles y aquellos con índices de discriminación inferiores a .3 (Nunnally, 1972). En total, se obtuvieron 128 ítems con buen desempeño. Finalmente, en la séptima etapa se incorporó la aleatorización de los distractores, en que el sistema automáticamente generaría distractores

diferentes para cada estudiante, cada vez que rindiera la prueba. La aleatorización debía cumplir con los criterios estipulados respecto del nivel de dificultad de la palabra, categoría gramatical y capacidad de ser graficada. En esta etapa, realizamos un segundo pilotaje con más de 1.600 estudiantes de educación parvularia y básica hasta octavo año, de nivel socioeconómico alto, medio alto, medio, medio bajo y bajo, para comparar desempeño en ítems con distractores fijos y aleatorios en las dos nuevas formas.

Dado que el tamaño del *corpus* lingüístico de textos escolares chilenos era relativamente pequeño, quisimos fortalecer la validez del *corpus* comparándolo con un *corpus* de 30 millones de palabras en español. La correlación entre palabra base y frecuencias fue de .78, lo que sugiere que las palabras base ocurrían con similar frecuencia en ambos *corpus*, y que el uso léxico en estos dos *corpus* era bastante similar.

Participantes

En el primer piloto participaron 571 alumnos de cuatro establecimientos (uno municipal, dos subvencionados y uno privado), y de los niveles kínder a cuarto año básico, representativos de la población escolar chilena. En el segundo pilotaje, participaron 1.660 alumnos de 6 establecimientos (cuatro municipales, uno subvencionado y uno privado) de kínder a octavo básico. 294 alumnos provenían del colegio subvencionado, 539 del privado y 827 de los municipales. La distribución de los estudiantes por curso se detalla en las tablas 1 y 2. De estos alumnos 1.060 respondieron la forma A y 600 respondieron la forma B.

Instrumentos

La prueba EVOC consta de dos formas paralelas y se aplica mediante computador, salvo con los alumnos de kínder, a quienes se les aplica en una *tablet*, por ser un formato más fácil para su edad. Los alumnos reciben las instrucciones verbales de la prueba por audio, por lo que cada computador o cada *tablet* tienen audífonos que permiten a los alumnos tomar la prueba en paralelo sin interrupciones ni ruido ambiente. Se incluyen dos ejemplos para la ejercitación antes de iniciar la prueba. En la pantalla, las imágenes no se activan hasta cuando el alumno haya escuchado la palabra, para evitar que responda sin poner atención al estímulo. Una vez que los alumnos han escuchado la palabra clave, las cuatro imágenes (alternativas de respuesta correcta) de la pantalla se iluminan y solo entonces el alumno puede marcar su respuesta. En los dos estudios piloto, todos los estudiantes respondieron todas las preguntas.

Tabla 1
Participantes por estudio, según curso y tipo de colegio

Curso	Piloto 1 (N = 571)			Piloto 2 (N = 1.661)		
	Municipales	Privados	Subvencionados	Municipales	Privados	Subvencionados
Kínder	25	15	74	74	62	57
Primero	25	15	75	104	80	58
Segundo	19	15	77	90	56	61
Tercero	24	15	75	84	59	58
Cuarto	28	15	74	90	61	59
Quinto	-	-	-	87	47	0
Sexto	-	-	-	96	53	0
Séptimo	-	-	-	117	49	0
Octavo	-	-	-	86	73	0
Total	121	75	375	827	539	294

Fuente: elaboración propia

Tabla 2
Total de participantes por estudio según curso

Curso	Piloto 1 (N = 571)	Piloto 2 (N = 1.661)
Kínder	114	93
Primero	114	242
Segundo	111	207
Tercero	114	201
Cuarto	118	210
Quinto	-	134
Sexto	-	149
Séptimo	-	166
Octavo	-	159
Total	571	1.661

Fuente: elaboración propia

Resultados

Estudio piloto 1

De los 571 alumnos evaluados, 289 y 282 tomaron las formas A y B, respectivamente y representan los distintos niveles socioeconómicos de la población escolar chilena. El valor promedio de las respuestas correctas en este piloto para la muestra completa (N = 571) fue de 44.23 (DS 12.25) con un rango de 14 a 80 respuestas correctas, de un máximo

posible de 156. Como era esperable, los puntajes promedio son más altos en los cursos superiores que en los inferiores. La desviación estándar se incrementó también en los cursos superiores en la forma A, pero disminuye en los cursos superiores en la forma B. La distribución de los puntajes en las formas A y B se acerca a la normalidad (Forma A: skew = -.50, kurtosis = 2.81; Forma B: skew = .20, kurtosis = 2.22). Los resultados por curso y por forma se presentan en la tabla 3. Tal como se aprecia, la media general para la forma B fue superior a la de la forma A en todos los cursos, lo que sugiere que la forma B tendría un nivel de dificultad inferior. Ante este resultado, se estimó conveniente reorganizar los ítems de las dos formas para asegurar niveles comparables de dificultad en ambas.

Tabla 3

Medias (SD) y rangos para los cursos kínder a cuarto básico. Estudio piloto 1

Nivel educativo	Forma A			Forma B		
	n	Media(SD)	Rango	n	Media (SD)	Rango
Total	289	61.16 (12.25)	31-92	282	68.80 (11.60)	34-94
Kínder	57	53.37 (7.84)	35-69	57	54.86 (8.62)	34-68
1	57	56.47 (9.93)	32-72	57	62.91 (8.19)	44-75
2	59	61.31 (11.64)	42-83	52	71.15 (7.31)	56-91
3	57	65.16 (10.78)	31-84	57	75.19 (5.51)	58-89
4	59	69.20 (13.43)	44-92	59	79.69 (6.98)	50-94

Fuente: elaboración propia

Análisis de los ítems

Para cada forma, revisamos la dificultad de los ítems. Las palabras se ordenaron según frecuencia, considerando que ella es un *proxy* de familiaridad léxica y, por ende, de dificultad (Chen & Meurers, 2016; Ryder & Slater, 1988). Se consideró como “difícil” un ítem, si el puntaje obtenido era inferior a .3; “aceptable”, si era superior a .3 e inferior a .7; y “fácil”, si el puntaje obtenido era superior a .7. Los niveles de dificultad resultaron ser bastante variados en ambas formas.

Se analizó el índice de discriminación de cada ítem, para fortalecer la validez del instrumento y garantizar que el ítem midiera efectivamente el constructo de vocabulario. Se utilizaron dos métodos para analizar el índice de discriminación: 1) se consideraron las correlaciones punto-biserales y, 2) el índice d. Se consideraron aceptables correlaciones punto biserales de .2 o más. Para el índice d ($d = (H-L/N)*2$) (Karelia, Pillai & Vegada, 2013), en el que el grupo con menor desempeño corresponde al 27% de los alumnos

en el total de respuestas correctas al instrumento (Kelley, 1939), y el “grupo con más alto desempeño” corresponde al 27% de los puntajes más altos; se consideraron *bajos* los índices inferiores a .2, *aceptables* aquellos entre .2 y .24, *buenos* aquellos entre .25 y .35 y *excelentes* aquellos por sobre el .35. Hubo 23 y 18 ítems de las formas A y B con índices bajos. Con estos resultados, se crearon las nuevas formas A y B con 64 ítems cada una.

Validez

Se revisó la validez concurrente de EVOC en relación con la prueba TEVI-R (Echeverría, Herrera & Segure, 2002). A una submuestra de 403 estudiantes ($n = 208$ para la forma A y $n = 195$ para la forma B), además de EVOC, se le administró la prueba TEVI. Para validar EVOC, eliminamos los ítems con bajo índice de discriminación, e implementamos un “techo” de 8 respuestas incorrectas seguidas para ajustar EVOC a los procedimientos de administración del TEVI. Los resultados de EVOC y TEVI mostraron una distribución normal, con una correlación pequeña pero positiva y significativa ($r = .133$, $p < .001$) entre los puntajes de ambas evaluaciones. Pensamos que la baja correlación se puede deber a la ausencia de medidas basales en EVOC, el cual, a diferencia del TEVI, solo contaba con medida de “techo”. También se eliminaron los puntajes 0 en el TEVI ($n = 4$ Forma A, $n = 2$ Forma B). Para ambas formas, la correlación de Pearson fue pequeña pero positiva y significativa ($r = .37^*$ Forma A, $r = .34^{**}$ Forma B), con lo que ambas formas demostraron validez frente al TEVI.

Confiabilidad

La confiabilidad del instrumento se determinó analizando su consistencia interna. El índice Split-half, que divide los puntajes de cada ítem en dos submuestras y las correlaciona, entregó un alfa de .86, lo que indica buena consistencia interna.

Análisis psicométricos

Se usó el modelamiento Rasch estándar para la construcción psicométrica de instrumentos de evaluación. Se usó el modelo Rasch, por una parte, para poder ubicar a los estudiantes y a los ítems en una misma escala, y por otra, identificar qué ítems demostraban una falta de calce entre el modelo y los datos. En la tabla 4 se presenta la dificultad Rasch, el *infit* y el *outfit*, las medidas de punto y el índice de discriminación entre el 27% superior e inferior de los evaluados. Los ítems fueron considerados como buenos si cumplían al menos dos de los siguientes criterios: un estadístico de ajuste (*infit*) de entre .5 y 1.5; un estadístico de ajuste para casos atípicos (*outfit*) de entre .5 y 2.0, y punto biserial e índice de discriminación de .2 o más (Bond

& Fox, 2015; Linacre, 2002). El resultado del análisis Rasch fue una prueba confiable, con una validez limitada, y verticalmente alineada para los niveles de kínder a cuarto año básico. Tal como se muestra en la figura 2, los ítems correspondientes a palabras de mayor frecuencia parecen ser más fáciles para los alumnos, mientras que las palabras de menor frecuencia pueden ser tanto fáciles como difíciles.

Tabla 4

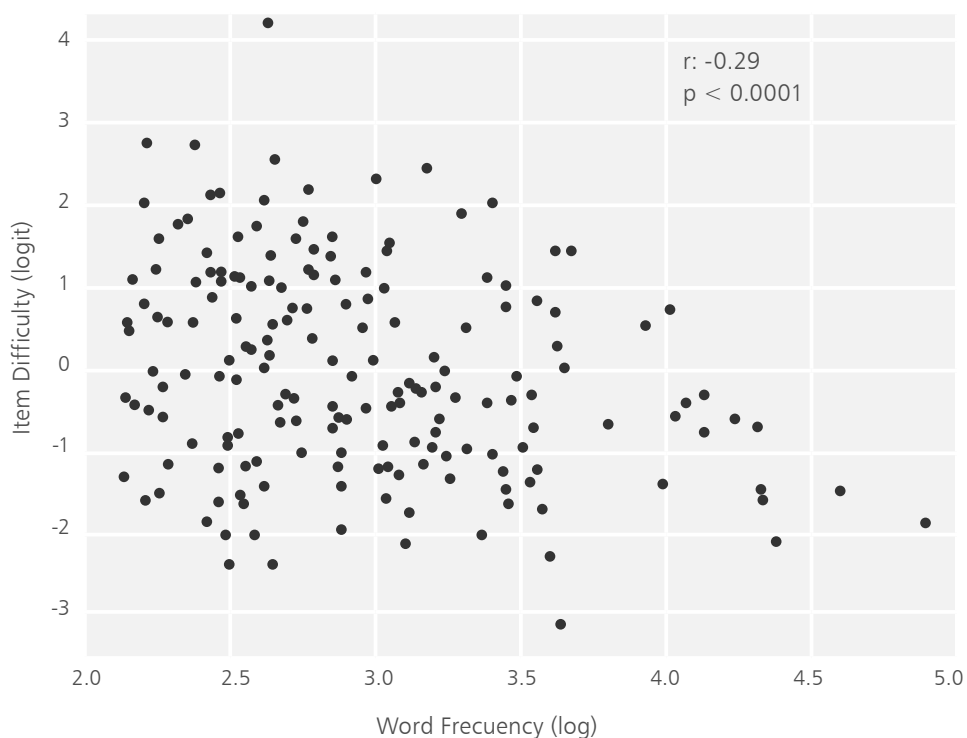
Medias (SD) y rangos estudio piloto 1 (kínder a cuarto básico)

Nivel educativo	Forma A			Forma B		
	n	Media (SD)	Rango	n	Media (SD)	Rango
General	289	61.16 (12.25)	31-92	282	68.80 (11.60)	34-94
Kínder	57	53.37 (7.84)	35-69	57	54.86 (8.62)	34-68
1	57	56.47 (9.93)	32-72	57	62.91 (8.19)	44-75
2	59	61.31 (11.64)	42-83	52	71.15 (7.31)	56-91
3	57	65.16 (10.78)	31-84	57	75.19 (5.51)	58-89
4	59	69.20 (13.43)	44-92	59	79.69 (6.98)	50-94

Fuente: elaboración propia

Figura 2

Comparación de frecuencia de palabras y dificultad en el piloto 2



Fuente: elaboración propia

Estudio piloto 2

En el segundo estudio piloto, reclutamos 1.661 estudiantes de kínder a octavo año básico, de los tres tipos de dependencia. Los alumnos fueron evaluados con EVOC al inicio del segundo semestre. Las instrucciones, procedimientos y formato fueron los mismos que los realizados en el primer piloto, salvo que los alumnos debieron completar los 64 ítems de la forma que se les asignó (A o B). Los alumnos que no completaron todos los ítems fueron excluidos del análisis. 1.013 alumnos respondieron la forma A completa y 638 la forma B completa.

La tabla 5 muestra las medias, desviaciones estándar y rangos para los totales de cada forma en los distintos cursos. Tal como se esperaba, los puntajes promedios en los cursos inferiores son menores que los de los alumnos en los cursos superiores, y se observó mayor variabilidad de los puntajes promedio en los cursos que van de quinto a octavo básico. Los análisis de varianza mostraron diferencias estadísticamente significativas para los puntajes promedio de cada curso $F(8, 1013) = 51.981, p < .000$, $F(8, 638)$. La confiabilidad interítem fue de .931, lo que indica alta consistencia interna.

Tabla 5

Estadísticas descriptivas. Segundo estudio piloto

Curso	Forma A		Forma B	
	M(SD)	Rango	M(SD)	Rango
Kínder	30.5 (7.19)	2-46	36.3 (9.7)	14-51
Primero	32.1 (10.90)	2-50	34.5 (13.1)	1-53
Segundo	39.2 (9.74)	2-54	44.4 (10.6)	6-56
Tercero	42.5 (8.84)	5-56	47.5 (10.1)	11-57
Cuarto	47.3 (7.20)	7-57	52.2 (5.4)	30-61
Quinto	47.1 (11.60)	5-60	49.5 (11.9)	1-60
Sexto	48.7 (13.80)	3-59	51.4 (8.8)	6-61
Séptimo	49.1 (12.90)	1-62	51.7 (10.5)	13-61
Octavo	43.4 (1-59)	1-59	48.2 (15.7)	5-62

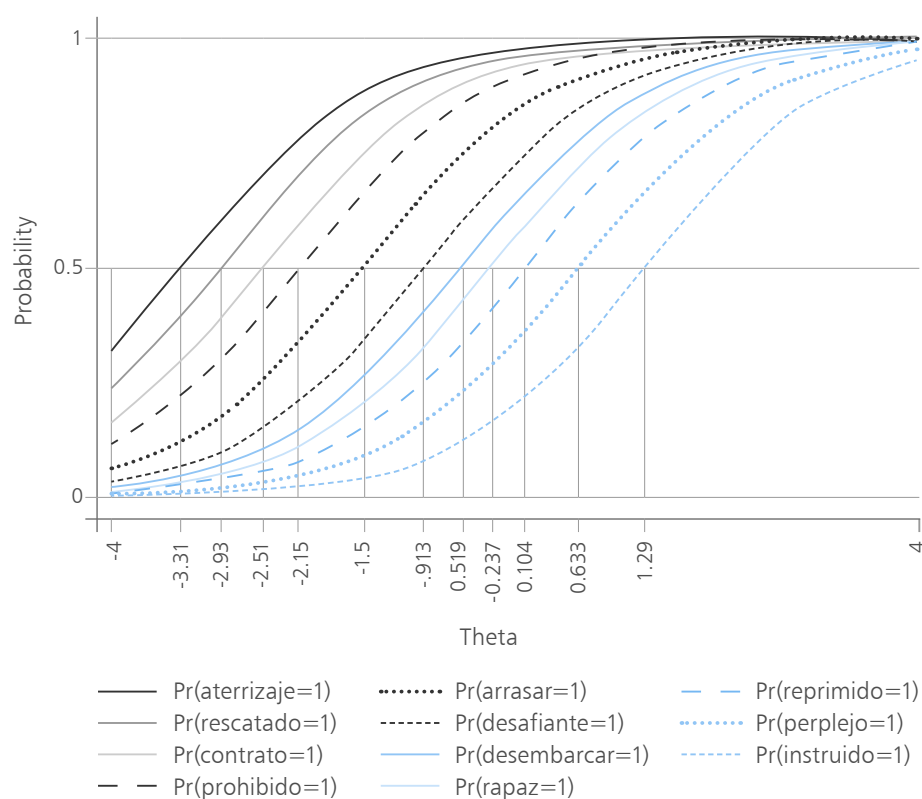
Fuente: elaboración propia

Si se comparan los promedios de ambas formas, los obtenidos en la forma B son notoriamente más altos que los de la Forma A en varios de los cursos, lo que podría indicar que la Forma B es más fácil para los estudiantes que la Forma A. Estas diferencias fueron estadísticamente significativas, según ANOVA, para kínder, segundo, tercero y cuarto básico. Al

graficar las distribuciones de puntajes para ambas formas, vemos que para la Forma A hay un importante número de casos que están por debajo del rango intercuartil en los estudiantes de quinto, sexto, séptimo y octavo, lo que se refleja en un desempeño inferior del grupo que respondió la Forma A, que además casi duplicaba en número al que respondió la Forma B. Para saber si la aleatoriedad en la presentación de los distractores incide en el desempeño de los estudiantes en la prueba EVOC, se analizaron las curvas de características de los ítems y sus correspondientes grados de dificultad. La figura 3 muestra los valores de dificultad por ítem en el eje X, y las líneas verticales de referencia muestran el lugar en el que este nivel equivale al nivel de dificultad del ítem al cruzar la curva en una probabilidad de .5. Esta figura corresponde a los ítems con distractores aleatorios. Por ejemplo, la línea al extremo izquierdo muestra que la palabra “aterri-zaje” respondida correctamente con una probabilidad de .5, y lo que da un resultado de -3.31, lo que quiere decir que el ítem es relativamente fácil dado que la característica asociada a un nivel probabilístico de .5 es inferior a la de otros ítems.

Figura 3

Características de los ítems con distractores aleatorios (muestra)

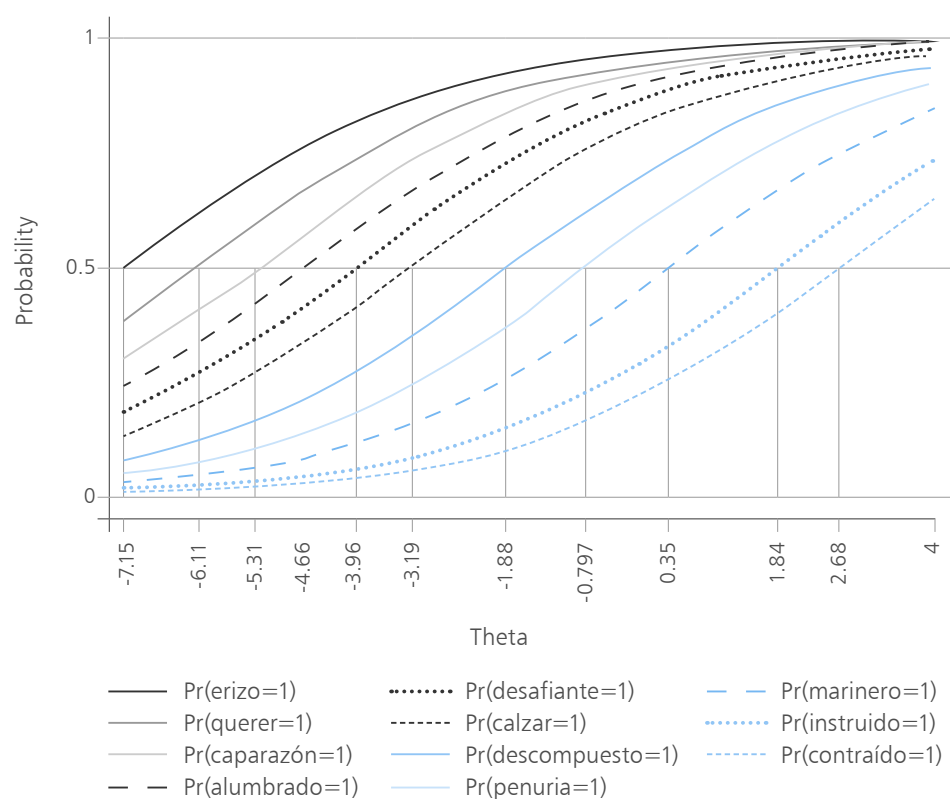


Fuente: elaboración propia

En el caso de los ítems con distractores fijos, en cambio, se observa mayor expansión en los niveles de dificultad de los ítems, que van de 7.15 a 2.679 (figura 4). Un aspecto interesante resultó ser que palabras como “erizo” resultaron ser fáciles tanto cuando los distractores eran fijos como cuando eran aleatorios. Lo mismo ocurrió con la palabra “contraído” que resultó ser difícil tanto con distractores fijos como aleatorios, lo que, según el modelo Rasch, ratifica la noción de que los ítems mantienen su dificultad en distintos contextos o situaciones evaluativas, o de que generan mediciones que no dependen de las muestras. Esta característica puede analizarse de manera más sencilla al correlacionar las dificultades de los ítems aleatorios y fijos. La correlación entre los puntajes del piloto 1 y del piloto 2 fue positiva, $r = .728$, $p < .05$. Además, se identificó una correlación positiva y significativa entre los errores estándar de los ítems con distractores fijos y aleatorios en ambos estudios, $r = .558$, $p < .05$. Las figuras 5 y 6 muestran el comportamiento del error estándar en ambas situaciones.

Figura 4

Características de los ítems con distractores fijos



Fuente: elaboración propia

Figura 5

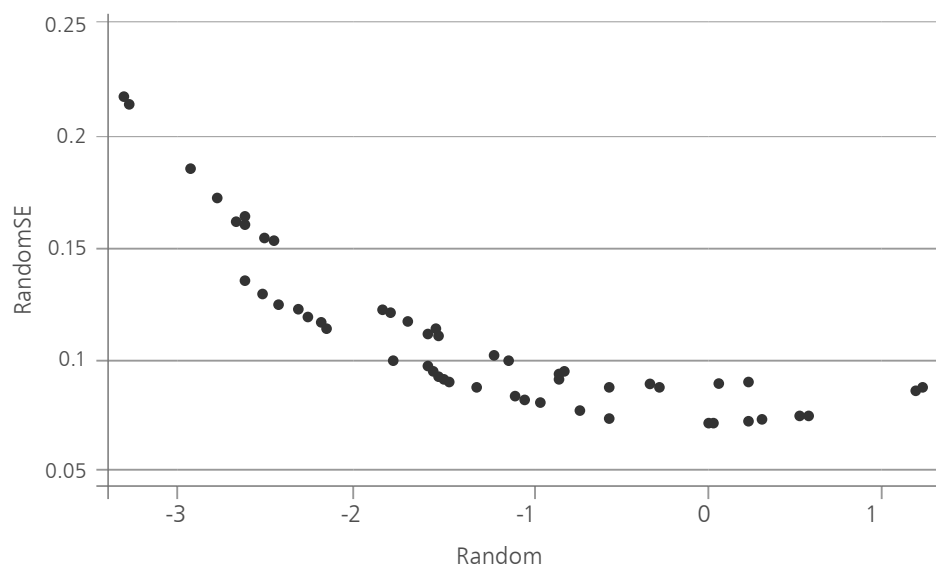
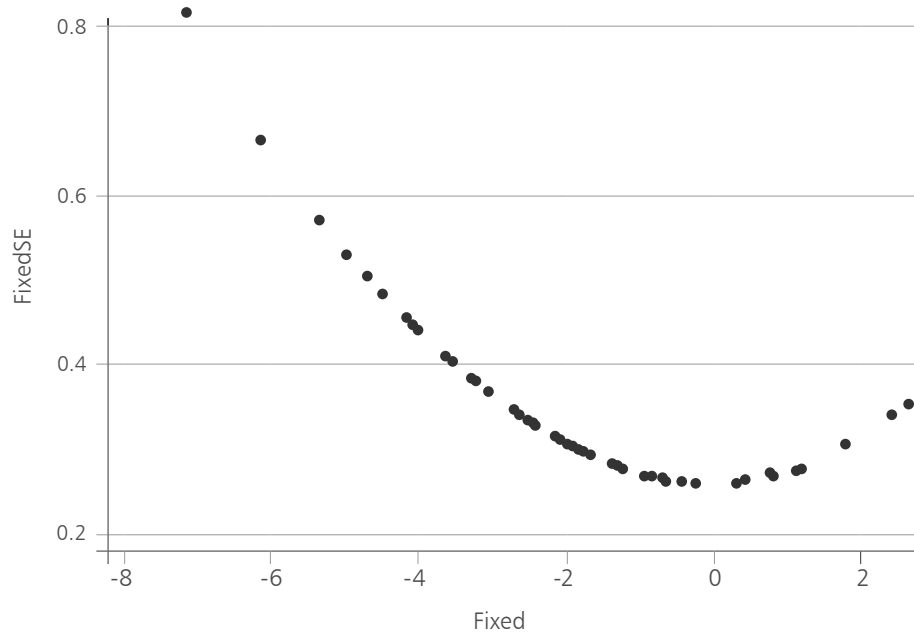
Dificultades según error estándar. Distractores aleatorios*Fuente: elaboración propia*

Figura 6

Dificultades según error estándar. Distractores fijos*Fuente: elaboración propia*

Como se observa especialmente en los tramos inferiores de los dos gráficos, a medida que las dificultades son más extremas (en el rango negativo, es decir, ítems demasiado fáciles), también aumenta el error

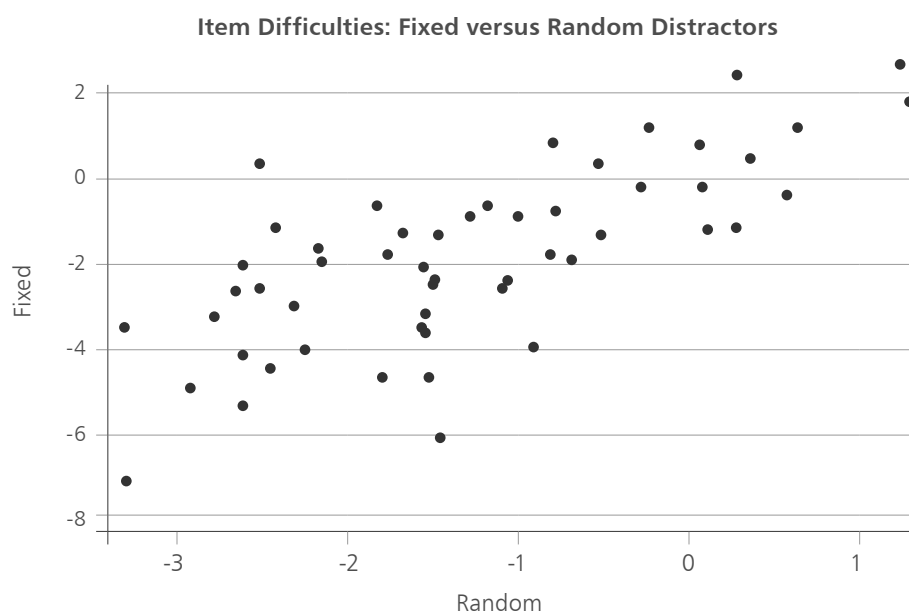
estándar, lo que resulta esperable. También es esperable que los errores estándar tengan menor correlación que la dificultad de los ítems, dado que probablemente dependen de factores externos a los ítems con distractores fijos.

Finalmente, en la figura 7 es posible apreciar la dispersión de los resultados para ítems con distractores fijos y aleatorios en el segundo estudio piloto. El hecho de que el rango es mayor para los ítems con distractores fijos es consistente con la interpretación del modelo Rasch. La dificultad del ítem depende en parte de la palabra que el alumno debe identificar y en parte, de los distractores. En la condición con distractores fijos, hay mayor probabilidad de que ocurra un valor extremo, por ejemplo, cuando una palabra difícil aparece con distractores difíciles. En la condición aleatoria, la contribución de los distractores a la dificultad del ítem se minimiza debido a que se dispersa en la muestra.

Figura 7

Dispersión de resultados ítems con distractores fijos y aleatorios.

Segundo estudio piloto



Fuente: elaboración propia

Discusión

Tras el pilotaje de la prueba EVOC, los resultados dan cuenta de que el instrumento permite evaluar el vocabulario receptivo de estudiantes chilenos de kínder a octavo básico de manera válida y confiable. El instrumento

cuenta con las características psicométricas y pedagógicas que lo hacen útil para determinar hasta qué punto los estudiantes conocen el significado de palabras de distinta frecuencia y dificultad, pero de bastante cotidianeidad. En línea con lo sostenido por P. David Pearson, Elfrieda H. Hiebert y Michael L. Kamil (2007) e Isabel Beck y Margaret McKeown (1991), EVOC es un instrumento construido en función de las palabras de nivel 2; es decir, aquellas que se encuentran en las experiencias diarias de lectura. En cierto modo, estas palabras no solo funcionan adecuadamente como una evaluación del conocimiento específico de los lectores del significado de las palabras en el contexto en que se encuentran, sino también como una medida de comprensión del texto en sí; es un vocabulario correspondiente al discurso académico, por lo que es propio de la enseñanza en la escuela. Se trata del vocabulario dominado por los lectores más competentes, de palabras que corresponden generalmente a conceptos familiares y ampliamente comprendidos, pese a que las palabras en sí no sean tan familiares a los lectores menos experimentados. También se trata de palabras que son necesarias para comprender al menos una parte local del contexto y que se relacionan con ideas esenciales del texto, pues, de no entender el significado de ellas podría verse amenazada la comprensión de las ideas textuales. Finalmente, se trata de términos que los estudiantes de seguro encontrarán en los textos correspondientes a su nivel de enseñanza, o como las denominan Isabel Beck, Margaret McKeown y Linda Kucan (2002), palabras poco comunes para conceptos relativamente comunes.

Específicamente, es importante destacar que el *corpus* de palabras de la prueba EVOC ha sido construido a partir de los textos escolares, lecturas complementarias y algunas colecciones de bibliotecas de aula de los estudiantes chilenos de kínder a octavo año básico. Se trata, por tanto, de un *corpus* que contiene todas las palabras que los estudiantes encontrarían en las asignaturas de lenguaje y comunicación, ciencias naturales y ciencias sociales y lectura complementaria a lo largo de la enseñanza básica. Esto aseguraría la exposición de los estudiantes a ellas, y por lo mismo, su posible comprensión.

Desde el punto de vista teórico, EVOC se fundamenta en la relación consistente y empíricamente validada entre comprensión lectora y vocabulario. Las relaciones entre estos constructos se han explicado de diferentes modos: por ejemplo, la hipótesis instrumentalista establece una relación causal por la cual conocer las palabras produce mejor comprensión lectora, mientras que la hipótesis de conocimiento del mundo sostiene que tanto la comprensión como el vocabulario son el producto de un conocimiento conceptual mayor. EVOC es un instrumento que evalúa el conocimiento del significado de las palabras independientemente de los distractores. La fuerte

correlación entre dificultad de ítem y errores estándar para distractores fijos y aleatorios indican que incluso al cambiar los distractores de esta forma las palabras que los niños deben reconocer no varían en su ordenación según dificultad léxica. Esto evidencia la capacidad del instrumento para evaluar el conocimiento de las palabras por parte de los estudiantes, dado que la aleatorización de distractores no fue un factor que incidió en su capacidad de reconocer su significado.

Pese a lo anterior, es importante establecer que los análisis realizados y presentados en este estudio establecen una validez limitada de EVOC, por cuanto se refieren a validez concurrente, establecida al comparar el desempeño de los estudiantes en relación con la prueba TEVI-R, de similares características. En estudios posteriores será posible determinar su validez predictiva. Esto podría realizarse, por ejemplo, midiendo el vocabulario receptivo de los estudiantes en kínder, para establecer si este predice la comprensión lectora posterior.

Conclusiones y limitaciones del estudio

El presente estudio da cuenta del diseño, pilotaje y validación limitada de EVOC, una prueba de vocabulario receptivo. EVOC es una prueba autoadministrada y digital, lo que permite conocer sus resultados de manera inmediata, y aplicarla en forma masiva, con un considerable ahorro de tiempo. Los análisis realizados dan cuenta de buena consistencia interna y validez concurrente, además de demostrar que la incorporación de distractores aleatorios validados por el modelo Rasch permite ubicar la dificultad del ítem y la habilidad léxica de la persona a lo largo de una misma escala, por lo que los ítems miden la misma característica. EVOC evalúa el conocimiento que los estudiantes tienen de las palabras, independientemente de los distractores que rodean la palabra que se evalúa.

Herramientas como EVOC son necesarias para que los docentes conozcan fehacientemente el grado de desarrollo de las habilidades léxicas de los estudiantes, toda vez que estas se relacionan tan estrechamente con la comprensión lectora. Por otra parte, y dado que EVOC fue construida a partir de los textos de kínder a octavo básico, esta herramienta permite saber si los alumnos podrán enfrentar esa lectura sin dificultades léxicas. Muchas de estas palabras constituyen lo que se denomina vocabulario académico, y que en su conjunto conforman un grupo de palabras necesarias para comprender, interpretar y utilizar textos en la vida escolar.

Una limitante importante de este estudio es que se cuenta con datos de solo dos estudios piloto, y pese a que se logró evaluar a un número importante de estudiantes, se requiere de más datos para ir calibrando los

ítems adecuadamente. El segundo piloto mostró diferencias en el grado de dificultad de ambas formas, lo que podrá ser ajustado en las sucesivas aplicaciones del instrumento. Otra limitación es el hecho de que en estos estudios se estableció una validez limitada, por cuanto se utilizaron correlaciones con otro test de características similares. La validez predictiva de EVOC debiera corroborarse mediante regresiones que permitan establecer si EVOC es un buen predictor de la comprensión lectora como se ha planteado desde la teoría.

Sobre los autores

Pelusa Orellana-García es profesora titular y vicedecana de investigación de la Facultad de Educación de la Universidad de los Andes, Chile. Experta en temas de lectura, en especial evaluación, diagnóstico y motivación.

María Francisca Valenzuela-Hasenohr es profesora asociada y vicedecana de investigación de la Facultad de Educación de la Universidad de los Andes, Chile. Experta en temas de desarrollo del lenguaje en primera infancia, en especial lenguaje oral, conciencia fonológica y vocabulario.

Melody Kung es posdoctora de la Universidad de Purdue, Indiana, Estados Unidos. Investigadora en educación con especialización en evaluación escolar.

Jeff Elmore es investigador, MetaMetrics, Inc., Estados Unidos.

A. Jackson Stenner es profesor e investigador, de la Universidad de North Carolina, Chapel Hill, Estados Unidos.

Referencias

- Baumann, J. F. & Kame'enui, E. J. (eds.) (2004). *Vocabulary Instruction: from Research to Practice*. New York: Guilford Press.
- Beck, I. & McKeown, M. (1991). Conditions of Vocabulary Acquisition. En M. L. Kamil, P. B. Mosenthal, P. D. Pearson & R. Barr (eds.). *Handbook of Reading Research*, volume II, 789-814. New York: Longman.
- Beck, I. & McKeown, M. (2007). Increasing Young Low-Income Children's Oral Vocabulary Repertoires through Rich and Focused Instruction. *The Elementary School Journal*, 107 (3), 251-271. DOI: 10.1086/511706
- Beck, I.; McKeown, M. & Kucan, L. (2002). *Bringing Words to Life. Robust Vocabulary Instruction*. New York: The Guilford Press.
- Beck, I.; Perfetti, C. A. & McKeown, M. G. (1982). Effects of Long-Term Vocabulary Instruction on Lexical Access and Reading Comprehension. *Journal of Educational Psychology*, 74 (4), 506-521. <http://dx.doi.org/10.1037/0022-0663.74.4.506>

- Biancarosa, G. & Griffiths, G. G. (2012). Technology Tools to Support Reading in the Digital Age. *The Future of Children*, 22 (2), 139-160. DOI: 10.1353/foc.2012.0014. Disponible en: <https://muse.jhu.edu/issue/27551>
- Biemiller, A. (2004). Teaching Vocabulary in the Primary Grades: Vocabulary Instruction Needed. En J. F. Baumann & E. J. Kame'enui (eds.). *Vocabulary Instruction: from Research to Practice*, 28-40. New York: Guilford Press.
- Biemiller, A. (2006). Vocabulary Development and Instruction: A Prerequisite for School Learning. En S. Neuman & D. K. Dickinson (eds.). *Handbook of Early Literacy Research*, volume 2, 41-51. New York: Guilford Press.
- Bond, T. G. & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. New York: Routledge.
- Cain, K. & Oakhill, J. (2011). Matthew Effects in Young Readers: Reading Comprehension and Reading Experience Aid Vocabulary Development. *Journal of Learning Disabilities*, 44 (5), 431-443. <https://doi.org/10.1177/0022219411410042>
- Chen, X. & Meurers, D. (2016). Characterizing Text Difficulty with Word Frequencies. En J. Tetreault, J. Burstein, C. Leacock & H. Yannakoudakis (eds.). *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications*, 84-94, San Diego, California, June 16, 2016. San Diego, California: Association for Computational Linguistics. Disponible en: <https://www.aclweb.org/anthology/volumes/W16-05/>, <https://www.aclweb.org/anthology/W16-0509>
- Darling-Hammond, L. & Bransford, J. (2005). *Preparing Teachers for a Changing World: What Teachers Should Learn and Be Able to Do*. San Francisco, California: Jossey-Bass.
- Dickinson, D. K. & Tabors, P. O. (2005). *Beginning Literacy with Language*. Baltimore: Paul H. Brookes Publishing.
- Dunn, L. M. (1986). TVIP: *Test de vocabulario en imágenes Peabody adaptación hispanoamericana* [Peabody Picture Vocabulary Test: Hispanic-American Adaptation]. Circle Pines, Minnesota: American Guidance Service.
- Echeverría, M. S.; Herrera, M. O. & Segure, J. T. (2002). *Test de Vocabulario en Imágenes TEVI-R*. Concepción, Chile: Ediciones Universidad de Concepción.
- Fien, H.; Santoro, L.; Baker, S. K.; Park, Y.; Chard, D. J.; Williams, S. & Haria, P. (2011). Enhancing Teacher Read Alouds with Small-Group Vocabulary Instruction for Students with Low Vocabulary in First-Grade Classrooms. *Small Psychology Review*, 40 (2), 307-318.
- Graesser, A. C.; McNamara, D. S.; Cai, Z.; Conley, M.; Li, H. & Pennebaker, J. (2014). Coh-Metrix Measures Text Characteristics at Multiple Levels of Language and Discourse. *The Elementary School Journal*, 115 (2), 210-229. <https://doi.org/10.1086/678293>
- Izura, C.; Cuetos, F. & Brysbaert, M. (2014). Lextale-Esp: A Test to Rapidly and Efficiently Assess the Spanish Vocabulary Size. *Psicológica*, 35, 49-66. Disponible en: <https://www.uv.es/revispsi/articulos1.14/3IZURA.pdf>
- Karelia, B. N.; Pillai, A. & Vegada, B. N. (2013). The Levels of Difficulty and Discrimination Indices and Relationship between them in Four-Response Type Multiple Choice Questions of Pharmacology Summative Tests of Year II MBBS [Bachelor of Medicine & Bachelor of Surgery] Students. *International e-Journal of Science, Medicine & Education, IeJSME*, 7 (2), 41-46.

- Disponible en: <https://pdfs.semanticscholar.org/e190/f7f97edae4eea6f979433d36430f34a597af.pdf>
- Kelley, T. L. (1939). The Selection of Upper and Lower Groups for the Validation of Test Items. *Journal of Educational Psychology*, 30 (1), 17-24.
- Linacre, J. M. (2002). What do Infit and Outfit, Mean-Square and Standardized Mean? *Rasch Measurement Transactions*, 16 (2), 878. Disponible en: <https://www.rasch.org/rmt/rmt162f.htm>
- Morris, A. (2011). Student Standardised Testing: Current Practices in OECD Countries and a Literature Review. *OECD Education Working Papers*, 65. OECD Publishing. Disponible en: <https://www.oecd-ilibrary.org/docserver/5kg3rp9qbnr6-en.pdf?expires=1564235500&id=id&accname=guest&checksum=881D15FD13AFF64A361BD240B7EEFE1C>
- Muñoz-Sandoval, A. F.; Woodcock, R. W.; McGrew, K. S. & Mather, N. (2005). *Batería III Woodcock-Muñoz: pruebas de aprovechamiento*. Rolling Meadows, Illinois: Riverside Publishing.
- Nagy, W. E. & Scott, J. (2000). Vocabulary Processes. En M. L. Kamil, P. B. Mosenthal, P. D. Pearson & R. Barr (eds.). *Handbook of Reading Research*, vol. III, 269-284. Mahwah, New Jersey: Erlbaum.
- National Assessment Governing Board, NAGB (2005). *Reading Framework for the 2009 National Assessment of Educational Progress*. Washington, D.C.: American Institutes for Research. Disponible en: <https://files.eric.ed.gov/fulltext/ED502953.pdf>
- National Institute of Child Health and Human Development, NICHD, National Institutes of Health, NIH (2000). *Report of the National Reading Panel. Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and its Implications for Reading Instruction*. Washington, D.C.: US Government Printing Office, NIH Publication No. 00-4769. Disponible en: <https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf>
- Neira, L. I. & Castro, F. (2008). Caracterización de los instrumentos de evaluación del desarrollo del lenguaje para hablantes del español. *Areté*, 8 (1), 53-62. Disponible en: <https://arete.iberu.edu.co/index.php/arete/article/view/435/400>
- Neuman, S. B. & Dwyer, J. (2009). Missing in Action: Vocabulary Instruction in Pre-K. *The Reading Teacher*, 62 (5), 384-392. Disponible en: <https://pdfs.semanticscholar.org/d3f0/effef6d7855f81ebef77ac754fe636882809.pdf>
- Nunnally, J. C. (1972). *Educational Measurement and Evaluation*. 2nd ed. New York: McGraw-Hill. Disponible en: <https://archive.org/details/educationalmeas000nunn>
- Organisation for Economic Co-operation and Development, OECD (2016). *Monitoring Quality in Early Childhood Education and Care Country Note: Chile*. Paris: OECD Publishing. Disponible en: <https://www.oecd.org/education/school/ECECDN-Chile.pdf>
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, UNESCO (2008). *Los aprendizajes de los estudiantes de América Latina y el Caribe. Primer reporte de los resultados del Segundo Estudio Regional*

- Comparativo y Explicativo, SERCE*. Santiago de Chile: UNESCO. Disponible en: <https://unesdoc.unesco.org/ark:/48223/pf0000160660>
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, UNESCO (2015). *Tercer Estudio Regional Comparativo y Explicativo, TERCE*. Santiago de Chile: Oficina regional. <https://unesdoc.unesco.org/ark:/48223/pf0000245060>
- Pearson, P. D.; Hiebert, E. H. & Kamil, M. L. (2007). Theory and Research into Practice: Vocabulary Assessment: What We Know and What We Need to Learn. *Reading Research Quarterly*, 42 (2), 282-296. <https://doi.org/10.1598/RRQ.42.2.4>
- Phythian-Sence, C. & Wagner, R. K. (2007). Vocabulary Acquisition: A Primer. En R. K. Wagner, A. E. Muse & K. R. Tannenbaum (eds.). *Vocabulary Acquisition: Implications for Reading Comprehension*, 1-14. New York: Guilford Press.
- Protopapas, A.; Sideridis, G. D.; Mouzaki, A. & Simos, P. G. (2007). Development of Lexical Mediation in the Relation between Reading Comprehension and Word Reading Skills in Greek. *Scientific Studies of Reading*, 11 (3), 165-197.
- Quinn, J. M.; Wagner, R. K.; Petscher, Y. & López, D. (2015). Developmental Relations between Vocabulary Knowledge and Reading Comprehension: A Latent Change Score Modeling Study. *Child Development*, 86 (1), 159-175. DOI: 10.1111/cdev.12292
- Rosas, R.; Ceric, F.; Aparicio, A.; Arango, P.; Arroyo, R.; Benavente, C.; Escobar, P.; Olguín, P.; Pizarro, M.; Ramírez, M. P.; Tenorio, M. & Véliz, S. (2015). ¿Pruebas tradicionales o evaluación invisible a través del juego?: Nuevas fronteras de la evaluación cognitiva. *Psykhé (Santiago)*, 24 (1), 1-11. Disponible en: http://www.psykhe.cl/index.php/psykhe/article/view/724/pdf_47
- Ryder, R. J. & Slater, W. J. (1988). The Relationship between Word Frequency and Word Knowledge. *The Journal of Educational Research*, 81 (5), 312-317. <https://doi.org/10.1080/00220671.1988.10885840>
- Schaefer, B.; Bowyer-Crane, C.; Herrmann, F. & Fricke, S. (2016). Development of a Tablet Application for the Screening of Receptive Vocabulary Skills in Multilingual Children: A Pilot Study. *Child Language Teaching and Therapy*, 32 (2), 179-191. <https://doi.org/10.1177/0265659015591634>
- Schatschneider, Ch.; Harrell, E. R. & Buck, J. (2007). An Individual-Differences Approach to the Study of Reading Comprehension. En R. K. Wagner, A. E. Muse & K. R. Tannenbaum (eds.). *Vocabulary Acquisition: Implications for Reading Comprehension*, 249-275. New York: Guilford Press.
- Shany, M. & Biemiller, A. (2010). Individual Differences in Reading Comprehension Gains from Assisted Reading Practice: Pre-Existing Conditions, Vocabulary Acquisition, and Amounts of Practice. *Reading and Writing: An Interdisciplinary Journal*, 23 (9), 1071-1083. DOI: 10.1007/s11145-009-9196-4
- Silverman, R. D. & Hartranft, A. M. (2015). *Developing Vocabulary and Oral Language in Young Children*. New York: The Guilford Press.
- Snow, C. E.; Burns, M. S. & Griffin, P. (eds.) (1998). *Preventing Reading Difficulties in Young Children*. Washington, D.C.: National Academy Press.
- Tannenbaum, K. R.; Torgesen, J. K. & Wagner, R. K. (2006). Relationships between Word Knowledge and Reading Comprehension in Third-Grade

- Children. *Scientific Studies of Reading*, 10 (4), 381-398. http://dx.doi.org/10.1207/s1532799xssr1004_3
- Thorne, C.; Morla, K.; Uccelli, P.; Nakano, T.; Mauchi, B.; Landeo, L.; Vásquez, A. & Huerta, R. (2013). Efecto de una plataforma virtual en comprensión de lectura y vocabulario: una alternativa para mejorar las capacidades lectoras en primaria. *Revista de Psicología (PUCP)*, 31 (1), 3-35. Disponible en: <http://revistas.pucp.edu.pe/index.php/psicologia/article/view/6369/6423>
- Verhoeven, L. & Leeuwe, J. van (2008). Prediction of the Development of Reading Comprehension: A Longitudinal Study. *Applied Cognitive Psychology*, 22 (3), 407-423. DOI: 10.1002/acp.1414
- Wagner, R.; Muse, A. & Tannenbaum, K. (2007). Promising Avenues for Better Understanding Implications of Vocabulary Development for Reading Comprehension. En R. K. Wagner, A. E. Muse & K. R. Tannenbaum (eds.). *Vocabulary Acquisition: Implications for Reading Comprehension*, 276-292. New York: Guilford Press.
- Walpole, S. & McKenna, M. C. (2007). *Differentiated Reading Instruction: Strategies for the Primary Grades*. New York: Guilford Press.
- Williamson, G. L.; Fitzgerald, J. & Stenner, A. J. (2014). Student Reading Growth Illuminates the Common Core Text-Complexity Standard: Raising Both Bars. *Elementary School Journal*, 115 (2), 230-254. DOI: 10.1086/678295