

---

# Interpretación edumétrica de los resultados de EVOC en estudiantes chilenos

## *An edumetric interpretation of EVOC test scores for Chilean students*

---

**PELUSA ORELLANA-GARCÍA**

Universidad de los Andes, Santiago de Chile  
porellan@uandes.cl  
<https://orcid.org/0000-0003-3851-8780>

**JOSÉ P. PEZOA**

Pontificia Universidad Católica de Chile  
jppezoa@uc.cl  
<https://orcid.org/0000-0002-4837-9289>

**Resumen:** El uso de resultados de mediciones de vocabulario ayuda a los docentes a monitorear los avances de sus estudiantes y a tomar decisiones pedagógicas para promover el aprendizaje. Este estudio tiene por objetivo adaptar la interpretación del instrumento EVOC, para medir amplitud de vocabulario, en 1.662 estudiantes de entre 6 y 14 años, a partir de bandas de frecuencia de palabras en un corpus. La metodología incluyó análisis factorial confirmatorio, la determinación de índices de dificultad empírica y el ensamblaje de los ítems en términos de frecuencia de aparición de las palabras en el corpus de textos escolares. Los resultados muestran que es posible analizar puntajes a partir de un estándar de desarrollo de la competencia léxica, aprovechando mejor esta información en la enseñanza y aprendizaje de lectura y vocabulario.

**Palabras clave:** Vocabulario, Comprensión lectora, Currículo escolar, Formatos digitales, Significado, Lectura.

**Abstract:** The use of vocabulary assessment scores can help teachers monitor student progress and make pedagogical decisions that promote learning. The purpose of the study was to adapt the interpretation of scores yielded by EVOC, a Spanish vocabulary assessment tool, to measure vocabulary range among 1,662 students between the ages of 6 and 14, based on word frequency bands from a corpus. The methodology included confirmatory factor analysis, the establishment of empirical difficulty indices, and the assemble of the items in terms of the order of appearance of the words in the corpus of school texts. The results show that it is possible to interpret scores in terms of lexical competence development standards, which should provide for better use of data in the teaching and learning of reading and vocabulary.

**Keywords:** Vocabulary, Reading comprehension, School curriculum, Digital formats, Meaning, Reading.

## INTRODUCCIÓN

Conocer el significado de una palabra es necesario para entender cualquier texto. Un vocabulario robusto permite acceder a recursos semánticos importantes para relacionar ideas y comprender a niveles más allá de lo literal (Kamil y Hiebert, 2005). Los lectores que conocen un amplio rango de palabras, sus acepciones y usos, tienen una ventaja importante respecto a lectores menos competentes para asociar conocimientos previos con ideas nuevas en un texto (Rasinski y Rupley, 2019; Wright y Cervetti, 2017). Tienen, además, mayor capacidad de comprender los matices de significado que una determinada palabra comporta en un contexto específico, lo cual es especialmente necesario en la lectura literaria. Por otra parte, una de las principales experiencias vicarias para el incremento del vocabulario es la lectura extensa e independiente (Duff, Tomblin y Catts, 2015), en la cual los lectores se enfrentan a un número importante de palabras desconocidas. De hecho, se estima que alrededor del 75% de las palabras nuevas que incorporamos a nuestro léxico provienen de lo que leemos. Tanto desde la investigación como desde la práctica pedagógica, el vocabulario y su enseñanza son aspectos centrales en la enseñanza, especialmente si se trata de mejorar la comprensión lectora de los estudiantes (Hiebert, Scott, Castaneda y Spichtig, 2019).

La relación entre comprensión y vocabulario ha sido ampliamente estudiada en la literatura. Se estima que su correlación está en un rango de .3 a .8 (Pearson, Hiebert y Kamil, 2007; Tannenbaum, Torgesen y Wagner, 2006), y que ésta se fortalece con la edad (Torgesen, Wagner y Rashotte, 1997). Considerando la magnitud y significancia de esta correlación, muchas intervenciones para mejorar las habilidades de comprensión a nivel escolar incluyen estrategias para estimular e incrementar el conocimiento de palabras que los estudiantes podrían encontrar en los textos que leen (Nagy, Anderson y Herman, 1987).

En el marco teórico, primero se presentan y discuten distintas miradas teóricas acerca del vocabulario y se profundiza acerca de su relación con la comprensión lectora. En segundo lugar, se revisan los aspectos del vocabulario a considerar para la construcción de instrumentos. A continuación, se aborda la utilización de corpus lingüístico en la medición de vocabulario. Se explican además las características del corpus utilizado en la elaboración de la prueba EVOC. Posteriormente, se profundiza en la relación entre frecuencia y dificultad en el corpus. Finalmente, se argumenta el uso de EVOC como alternativa a otros instrumentos para la evaluación del vocabulario en español, y se propone una interpretación de sus resultados desde un enfoque edumétrico.

## CONCEPCIONES TEÓRICAS ACERCA DEL VOCABULARIO

Existe gran consenso en relación al papel que cumple el vocabulario en el aprendizaje de lengua, en la capacidad de expresarse, de manejar conceptos, y en la adquisición de un segundo idioma (Medellín y Rodríguez, 2014). El vocabulario de una persona se define como la cantidad de unidades léxicas individuales cuyo significado conoce (Read, 2004), también conocido como vocabulario receptivo. El vocabulario expresivo, en cambio, es el conjunto de palabras que la persona, además de conocer su significado, es capaz de usar en sus interacciones cotidianas. También se ha afirmado que el vocabulario es un constructo multidimensional, por cuanto lo que se define como “conocer una palabra” puede variar mucho según si se incluyen, por ejemplo, connotaciones, derivaciones, colocaciones, etc. (Hiebert *et al.*, 2019; Schmitt, 2014). Por lo tanto, la estimación del vocabulario de una persona depende de definiciones respecto a las unidades que se contabilizan. Daller, Milton y Treffers-Daller (2007) ilustran esta dificultad con un ejemplo sencillo, como es el conocimiento del significado de un verbo: trabajar. Si se conoce el significado del infinitivo ¿debe contabilizarse como una palabra? ¿Qué ocurre con todas las formas verbales, adverbios, sustantivos y adjetivos asociados a trabajar? Este es solo un ejemplo de las complejidades subyacentes a la concepción de lo que constituye el vocabulario y, en términos prácticos, las dificultades implícitas en su evaluación (Schmitt, 2014).

Este artículo se centra en el concepto de vocabulario receptivo; es decir, en aquellas palabras impresas que un lector encuentra, y a las cuales les asigna significado para lograr comprender un texto. Al enfrentarse a una palabra desconocida, el lector puede deducir su significado a partir del contexto, recurrir al diccionario, o bien descomponerla en partes conocidas, como prefijos, sufijos y raíces (Proctor, Silverman, Harring y Montecillo, 2012). Esta habilidad facilita la profundización en la interpretación de un texto y el uso de su información (e.g. Anderson y Pearson, 1984). Asimismo, conocer el significado de palabras asociadas a conceptos o contenidos disciplinares que un niño aprende en la escuela determina en gran medida el acceso al aprendizaje y al capital cultural (Nation, 2006).

Anderson y Freebody (1981) proponen tres hipótesis para explicar la relación entre vocabulario y comprensión lectora. Una de ellas, la hipótesis instrumentalista, establece que el vocabulario es un factor causal y esencial para el desarrollo de la comprensión, que permite que exista buena comprensión de un texto. En la hipótesis de aptitud, en cambio, el vocabulario es más bien un resultado de una buena inteligencia verbal, y en la hipótesis que se basa en el conocimiento, el vocabulario es un indicador de que el lector posee una buena cantidad de conocimiento

del mundo y de conocimiento conceptual. Estas tres posibilidades sugieren que la relación entre ambos constructos puede ser tanto causal como de resultado, y que por tanto no se trata de una relación unidireccional ni de constructos unidimensionales.

Por otra parte, la lecturabilidad de un texto es la medida de dificultad que éste reporta para el lector, tanto en función de aspectos sintácticos como semánticos (Fry, 1968), y existen diferentes fórmulas cuantitativas para su medición. El aspecto semántico se mide en función de la frecuencia de aparición de las palabras (Chall, 1995), por lo que un texto con palabras de vocabulario menos frecuentes será más difícil de comprender que uno en el que aparecen palabras que son más comúnmente usadas. Esto refleja cómo el vocabulario y su frecuencia inciden en nuestra capacidad para comprender textos diversos.

#### LA EVALUACIÓN DEL VOCABULARIO

Una complejidad respecto de la medición del vocabulario radica en la distinción entre las dimensiones de amplitud, profundidad y fluidez de acceso (Proctor *et al.*, 2012; Qian, 2002). La amplitud se refiere a la cantidad de palabras que una persona conoce. Ha sido el aspecto más comúnmente evaluado del vocabulario, y aquello que mide el instrumento del presente artículo. La profundidad tiene que ver con el nivel de conocimiento que la persona tiene de dicha palabra y de su significado. Es un constructo más difícil de evaluar, y por tanto ha sido menos estudiado (Proctor *et al.*, 2012; Read, 2004). Finalmente, la fluidez de acceso se refiere a la capacidad de acceder y recuperar las unidades del léxico mental (Alber, 2016). De estos tres aspectos, muchos autores sugieren que la profundidad sería la que predice el desempeño lector, y estaría por tanto más directamente asociada a la comprensión lectora (Oullette, 2006; Schmitt, 2014) y con una especie de conciencia metalingüística en la cual confluyen aspectos semánticos, sintácticos y morfológicos del conocimiento de una palabra (Binder, Cote, Lee, Bessette y Vu, 2018; Proctor *et al.*, 2012). En el caso de la morfología, el conocimiento y el manejo de morfemas derivativos es el que más contribuye a mejoras en comprensión lectora (Kieffer y Lesaux, 2008). En relación con la semántica, su principal aporte tiene que ver con la polisemia; es decir, la toma de conciencia de que las palabras tienen múltiples significados asociados a contextos diferentes, y pareciera explicar diferencias significativas en incrementos de comprensión lectora en adolescentes más que en lectores principiantes (Oullette, 2006). Como han sugerido otros autores (e.g., Read, 2019), la estimación de la profundidad en el vocabulario puede medirse a través de aspectos de la conciencia metalingüística como el conocimiento morfológico, sintáctico

y semántico. Para medir conocimiento semántico se presentan tareas de sinonimia y polisemia (Oullette, 2006; Torgesen *et al.*, 2006), o de definiciones. En el caso de definiciones, el objetivo es poder determinar la calidad de la representación léxica que el sujeto tiene para la palabra.

#### SELECCIÓN MÚLTIPLE EN MEDICIONES DE VOCABULARIO

Existen pocos instrumentos con evidencia de validez y confiabilidad para la evaluación del vocabulario, y en muchos casos se requiere adaptarlos para responder a diferencias lingüísticas y culturales propias del español de distintos países. En el caso de instrumentos para evaluar vocabulario receptivo, el formato más comúnmente usado es el test con imágenes, en el que se pide al evaluado que identifique la imagen correspondiente al estímulo oral entregado (una palabra). Se presentan generalmente 4 imágenes, donde 1 corresponde al estímulo y 3 a distractores, siendo las cuatro de la misma categoría gramatical y de frecuencia similar. Algunos de estos instrumentos son la prueba Peabody en Imágenes (Dunn y Dunn, 1997), el TEVI y TEVI-R (Echeverría, Herrera y Segure 1995; 2002), y la Batería de Vocabulario de la prueba Woodcock-Muñoz (Schrank *et al.*, 2005). Estas pruebas se administran de forma individual, principalmente en contextos de consulta psicopedagógica, y en algunos casos requieren de una certificación especial para su uso, lo que los hace poco costo-efectivos.

Por otro lado, también existe preocupación ante la sobreestimación de puntajes debido a la adivinación de alternativas y medición de constructos irrelevantes en este formato de prueba (Gyllstad, Vilkaite y Schmitt, 2015). Por ejemplo, en pruebas de medición de vocabulario VLT, se ha encontrado que en promedio los puntajes aumentan un 16.7% debido al adivinamiento (Stewart y White, 2011). Otros estudios han encontrado que, al evaluar el tamaño del vocabulario, alrededor del 17% de las respuestas se responden correctamente al azar; es decir, sin que el evaluado conozca efectivamente su significado (Gyllstad *et al.*, 2015). Esto es más relevante cuando se considera que los evaluados contestan utilizando su conocimiento de los distractores más que del significado de la palabra, aumentando la probabilidad de contestar correctamente por azar (Stewart y White, 2011). Algunas estrategias para contestar las pruebas de alternativas son conocer efectivamente el significado de la palabra, inferirlo, eliminar alternativas poco plausibles, descartar opciones que el evaluado conoce, eliminar alternativas por asociación o agrupamiento de significados o adivinar ciegamente (e.g. Gyllstad *et al.*, 2015). Cuando el evaluado responde conociendo la palabra, la respuesta debería ser lo suficientemente rápida como para visualizar las alternativas, y responder sin nece-

sidad de inferir la respuesta correcta, lo que demoraría la velocidad de respuesta (Gyllstad *et al.*, 2015). Considerando que todas las demás estrategias de respuesta requieren procesamiento inferencial, lo que demora más tiempo que la respuesta directa del conocimiento de la palabra, se puede utilizar el tiempo de respuesta como mecanismo de control para saber qué ítems fueron respondidos a partir de la evocación del significado de la palabra y cuáles no.

Asimismo, las pruebas de vocabulario suelen ser demasiado extensas. Algunos han sugerido construir pruebas más cortas para evitar el agotamiento de los evaluados (Gyllstad *et al.*, 2015), lo que también permitiría incorporar ítems para testear en aplicaciones posteriores y continuar el proceso de mejora continua del instrumento.

#### EVALUACIÓN DE VOCABULARIO EN FUNCIÓN DE UN CORPUS Y BANDAS DE FRECUENCIA

Un corpus lingüístico es una colección de textos almacenados en una base de datos, usados generalmente para distintos tipos de análisis lingüísticos (Taylor y Baker, 2008). Los corpus contienen millones de palabras que provienen de cientos de miles de textos orales o escritos, y que representan el uso real de la lengua. Aunque originalmente un corpus incluía el lenguaje en general, en la actualidad se han elaborado corpus específicos que recopilan términos de distintos ámbitos, como, por ejemplo, corpus relacionado con la ciencia, o el español de una determinada región de Latinoamérica. El uso de corpus en el ámbito de la enseñanza del lenguaje o para fines de elaboración de instrumentos de medición de aspectos de la lengua no data de más de 20 años, y se centra en evaluaciones de aspectos sintácticos en un segundo idioma (Barker, Salamoura y Saville, 2015; Park, 2014; Simon, 2017). Su uso para investigar o evaluar la dimensión semántica de la lengua es mucho menos extenso (Anthony, 2017), pese a ser una herramienta valiosa para medir tanto amplitud como profundidad. Algunos estudios (e.g., Thompson y Sealey, 2007) han generado corpus usando como base compilaciones de lecturas literarias infantiles para comparar repertorios, identificar patrones lingüísticos en textos auténticos, etc.

Algunos expertos (e.g. Hiebert *et al.*, 2007; 2019) sugieren que los instrumentos diseñados se vinculen más directamente con los fines de dicha evaluación (si se va a medir vocabulario para la comprensión lectora, se debe evaluar vocabulario receptivo y no expresivo, por ejemplo). Otro aspecto fundamental es la selección de textos a partir de los cuales se elegirán las palabras incluidas en dichas evaluaciones, advirtiendo que distintos géneros textuales pueden incluir palabras

de mayor o menor dificultad. Textos disciplinares, de ciencia o matemática, por ejemplo, contienen términos específicos que son necesarios para comprender el contenido curricular, y que muchas veces no se evalúan (Hiebert *et al.*, 2019). Este estudio busca medir vocabulario receptivo de palabras propias del corpus escolar chileno, en diversas asignaturas cursadas entre kínder y octavo año básico (6 a 14 años).

En la evaluación educacional del vocabulario, las pruebas se construyen principalmente con una visión psicométrica que entrega la posición relativa del estudiante respecto de su muestra/población (TEVI, PPVT, EVOC 1, por ejemplo). Este enfoque es usado generalmente como evidencia para detectar algún problema específico de lectura, pero no da cuenta necesariamente de un estándar como la frecuencia de palabras. Por otro lado, los corpus de palabras usados han sido extraídos de diccionarios, y no responden necesariamente a la transacción de significado propio de la esfera escolar. Uno de los objetivos es adaptar el instrumento EVOC para ser interpretado en función de un estándar dado por la frecuencia de aparición de palabras en la esfera de la lectura escolar, y no desde una posición relativa a la muestra.

#### RELACIÓN ENTRE FRECUENCIA DEL CORPUS Y DIFICULTAD DEL ÍTEM

El corpus generado en el presente estudio incluye la revisión de 99 libros (textos escolares chilenos de educación preescolar y básica y otras lecturas escolares) de distintas asignaturas. Esta base textual generó alrededor de 3 millones de bases léxicas de palabras ordenadas según frecuencia de aparición. Considerando que el vocabulario se adquiere por procesos de interacción del evaluado con su ambiente próximo (e.g. Mendive, Mascareño Lara, Aldoney, Pérez y Pezoa, 2020; Rowe, 2012), se espera que, cuando exista mayor frecuencia de aparición de una palabra en el contexto, ésta sea más conocida (ítem más fácil). Por otro lado, los sistemas con los que interactúa el evaluado y, mediante los cuales adquiere vocabulario, se amplían desde kínder a octavo año básico (Bronfenbrenner y Morris, 2009), por lo que el entorno al cual pertenece el corpus recolectado es solo una fuente más de desarrollo. En este sentido, palabras que tengan una baja frecuencia en el corpus pueden haber sido aprendidas en otros contextos, como en la interacción con los padres. Por este motivo, se espera que palabras con mayor frecuencia de aparición sean en promedio más fáciles que palabras con menor frecuencia de aparición, pero no se espera que todas las palabras con mayor frecuencia sean más fáciles que todas las palabras con menor frecuencia de aparición.

## MATERIAL Y MÉTODOS

### *Objetivos*

Los objetivos del estudio son los siguientes:

- Adaptar el instrumento de evaluación de vocabulario EVOC (Orellana-García, Valenzuela-Hasenohr, Kung, Elmore y Stenner, 2020) para que pueda ser interpretado desde un punto de vista edumétrico y no puramente psicométrico.
- Modificar la interpretación de los resultados de EVOC a fin de que se refieran a un estándar o enfoque por competencias adquiridas, y no a partir de la distribución de la muestra y/o población.

### *Instrumentos*

EVOC 1 es una prueba de formas paralelas (formas C y D) que busca medir vocabulario receptivo y se aplica a través de plataforma digital, por medio de computadores o *tablets* (Orellana-García *et al.*, 2020). EVOC mide la amplitud del vocabulario a través de la frecuencia de bases léxicas o sus conjugaciones en el corpus de palabras que aparecen en textos escolares chilenos desde kínder (5 a 6 años) a octavo grado (14 años).

La prueba se aplica en una plataforma tecnológica que expande hacia computadores y *tablets* las formas en que tradicionalmente los estudiantes aprenden y se evalúan (Biancarosa y Griffiths, 2012; Pearson *et al.*, 2007). Cada computador o *tablet* tiene audífonos que permiten a los alumnos rendir la prueba en paralelo sin interrupciones ni ruido ambiente. La prueba consiste en la escucha de una palabra y la aparición de cuatro imágenes. El evaluado debe seleccionar la imagen que corresponde a la palabra escuchada. Las imágenes no se activan hasta que el alumno haya escuchado la palabra, para evitar que responda sin poner atención al estímulo. El objetivo de esta prueba es, principalmente, medir vocabulario receptivo. Para mayor información de la versión anterior del instrumento, se puede revisar Orellana-García *et al.* (2020).

### *Participantes*

Las formas C y D del instrumento EVOC fueron aplicadas a 1.023 y 639 estudiantes respectivamente. Estos participantes provienen de seis colegios, de los cuales

uno es particular pagado, tres colegios son particulares subvencionados y un colegio es municipal. En Chile, los establecimientos particulares pagados representan alrededor de un 5% del total de la matrícula escolar e incluyen a estudiantes de nivel socioeconómico alto. Por otra parte, un 50% de los estudiantes chilenos asiste a establecimientos particulares subvencionados y son de hogares de nivel socioeconómico medio; alrededor de un 45% asiste a establecimientos municipales y provienen de hogares de nivel socioeconómico medio-bajo o bajo (Centro de Estudios División de Planificación y Presupuesto, Mineduc, 2016). La proporción de estudiantes de cada uno de los tres tipos de dependencia escolar es similar en cada una de ellas (Tabla 1). La edad aproximada de los estudiantes se presenta en la Tabla 1.

**Tabla 1. Frecuencia y porcentaje de evaluados según nivel escolar**

NIVEL	EDAD	FRECUENCIA	%
Kínder	5-6	193	12%
Primero	6-7	242	15%
Segundo	7-8	207	12%
Tercero	8-9	201	12%
Cuarto	9-10	210	13%
Quinto	10-11	134	8%
Sexto	11-12	149	9%
Séptimo	12-13	166	10%
Octavo	13-14	159	10%

### *Análisis*

Para cumplir con el objetivo de adaptar la prueba EVOC 1 de modo tal que permita interpretar los resultados en términos de frecuencia de aparición de palabras en el corpus chileno, se siguieron tres pasos. En primer lugar, se realizó un análisis factorial confirmatorio de una dimensión latente (Conocimiento del vocabulario) utilizando el estimador WLSMV, que permite medir un constructo latente continuo a través de ítems binarios (Brown, 2015). Si bien para este tipo de análisis también es frecuente la utilización de modelos IRT, los parámetros de discriminación y dificultad utilizados en este modelo son análogos en ambos enfoques (Beaujean, 2015), pero en el enfoque factorial confirmatorio, las cargas factoriales pueden ser interpretadas en términos de la varianza del constructo latente. En este análisis se eliminan los ítems que no miden de forma adecuada el constructo latente.

En segundo lugar, y como la prueba está pensada para ser aplicada en distintos niveles educacionales, se espera que el ranking de dificultad de los ítems sea homogéneo a lo largo de los cursos. Esto quiere decir que los ítems más difíciles en un curso sean también los ítems más difíciles en otro curso. Para eso, se construyó el índice de dificultad empírica por curso y se estimó una matriz de correlaciones Spearman. Este coeficiente puede entenderse como el grado en que se mantiene el ordenamiento de los ítems en términos de ranking según dificultad (Field, Miles y Field, 2012). Se eliminaron los ítems de palabras que no presentan posiciones relativas similares en términos de dificultad a través de los diferentes cursos.

Como tercer paso, se ensamblaron los ítems de modo tal que los rangos de frecuencias de aparición de las palabras en el corpus escolar tuvieran una dificultad promedio menor que la dificultad mínima del siguiente nivel de frecuencia. Esto permite dar cuenta de que se espera que las palabras que aparecen con menos frecuencia en el entorno escolar del estudiante sean sustantivamente más difíciles. Finalmente, se estableció el criterio de corte de la prueba de modo tal que la probabilidad de seguir contestando la prueba por azar sea menor a la versión anterior.

## RESULTADOS

### *Casos perdidos*

La forma C fue aplicada a 1.023 niños o niñas, de los cuales 960 estudiantes tienen puntajes completos. La forma D fue aplicada a 639 niñas o niños, de los cuales 581 tienen puntajes completos. Se decidió trabajar mediante *listwise* debido a que en un análisis de validación es importante conocer el comportamiento empírico de los ítems. Además, la cantidad de valores perdidos es baja (3,1% en la forma C, y 4,5% en la forma D). La muestra total la constituyen 1.541 estudiantes.

### *Descriptivos*

Los estudiantes obtuvieron en promedio 45 respuestas correctas de las 64 que componen cada prueba. Se observa que, a medida que aumenta el nivel escolar al que pertenece el alumno, también aumenta la cantidad de respuestas correctas de los estudiantes, a excepción de octavo básico. Estos estudiantes obtienen puntajes más bajos que los niveles educacionales inferiores (Tabla 2). Esto puede deberse a que, al responder ítems que no presentan mucha dificultad, los alumnos de octavo básico dejan de contestar seriamente la prueba.

**Tabla 2. Cantidad de ítems respondidos según nivel escolar**

NIVEL ESCOLAR	N	RESPUESTAS CORRECTAS PROMEDIO	SD	MIN	MAX
Kínder	193	32	8.0	2	51
Primero	205	36	9.4	4	53
Segundo	197	42	8.4	6	56
Tercero	181	46	6.9	5	57
Cuarto	210	49	7.1	7	61
Quinto	125	50	9.3	1	60
Sexto	140	52	8.8	4	61
Séptimo	145	53	8.0	10	62
Octavo	145	49	15.4	3	62

*Análisis factorial confirmatorio y dificultad*

Se construyó un modelo de análisis factorial confirmatorio (AFC) para cada prueba utilizando el estimador WLSMV, que permite medir el constructo de vocabulario receptivo por medio de ítems ordinales binarios (Brown, 2015). En cada uno de los modelos se eliminaron los ítems que no son predichos por competencias narrativas con cargas factoriales de 0.4 o mayores, lo que es consistente con la recomendación de no interpretar cargas factoriales de 0.3 o menores (Harrington, 2015). En este procedimiento se eliminaron 18 ítems. Para determinar el modelo factorial final con ítems cuyo ranking de dificultad sea consistente a través de los cursos se estimó una matriz de correlaciones de Spearman, y se eliminaron 13 ítems cuya dificultad no era consistente a través de los cursos. Como se observa en la Tabla 3, los ítems seleccionados presentan ordenamientos de ranking de dificultad mayores .7 y en la mayoría de los casos mayores a .8.

**Tabla 3. Matriz de correlaciones Spearman forma C y D**

FORMA C

NIVEL ESCOLAR	KÍNDER	PRIMERO	SEGUNDO	TERCERO	CUARTO	QUINTO	SEXTO	SÉPTIMO	OCTAVO
Kínder	*****	0.957	0.945	0.92	0.906	0.7	0.854	0.817	0.785
Primero	<0.001	*****	0.968	0.953	0.941	0.754	0.887	0.871	0.799
Segundo	<0.001	<0.001	*****	0.974	0.951	0.78	0.894	0.883	0.827

[CONTINÚA EN LA PÁGINA SIGUIENTE]

**Tabla 3. Matriz de correlaciones Spearman forma C y D**

Tercero	<0.001	<0.001	<0.001	*****	0.964	0.806	0.92	0.906	0.854
Cuarto	<0.001	<0.001	<0.001	<0.001	*****	0.8	0.938	0.915	0.885
Quinto	<0.001	<0.001	<0.001	<0.001	<0.001	*****	0.813	0.84	0.774
Sexto	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	*****	0.912	0.884
Séptimo	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	*****	0.86
Octavo	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	*****

FORMA D

NIVEL ESCOLAR	KÍNDER	PRIMERO	SEGUNDO	TERCERO	CUARTO	QUINTO	SEXTO	SÉPTIMO	OCTAVO
Kínder	*****	0.925	0.881	0.884	0.893	0.866	0.775	0.79	0.746
Primero	<0.001	*****	0.949	0.896	0.883	0.889	0.82	0.784	0.75
Segundo	<0.001	<0.001	*****	0.933	0.884	0.9	0.864	0.84	0.803
Tercero	<0.001	<0.001	<0.001	*****	0.889	0.864	0.888	0.903	0.799
Cuarto	<0.001	<0.001	<0.001	<0.001	*****	0.869	0.829	0.821	0.739
Quinto	<0.001	<0.001	<0.001	<0.001	<0.001	*****	0.872	0.802	0.855
Sexto	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	*****	0.869	0.85
Séptimo	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	*****	0.808
Octavo	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	*****

Después de la selección de ítems en función del ranking de dificultad se re-estimaron los modelos factoriales (CFI=0.93; TLI=0.93; RMSEA=0.036; SRMR= 0.086 y CFI=0.91; TLI=0.91; RMSEA=0.033; SRMR=0.108) para los modelos C y D, respectivamente. De estos modelos factoriales confirmatorios se extrajeron los umbrales que se usaron en el ensamblaje.

*Ensamblaje según rangos y dificultad*

Los umbrales son un índice extraído del modelo factorial confirmatorio para respuestas binarias, análogo a la dificultad (*location*) en IRT, y representa el punto del nivel de habilidad a partir del cual es más probable que alguien que tiene cierto nivel de habilidad conteste correctamente el ítem (Beaujean, 2015). Al transformar este valor en la probabilidad acumulada de una curva normal, puede ser interpretado como la probabilidad de tener el ítem incorrecto (lo cual es diferente al concepto técnico de dificultad, que lo entiende como la probabilidad de tener el ítem

correcto). Este valor fue construido para el ensamblaje de los ítems, y analizado en conjunto con el ranking de aparición de palabras en el corpus. En la Tabla 4 puede observarse una columna que indica el ranking de aparición de palabras (frecuencia promedio en 100 libros) y una columna que indica la probabilidad de responder cada ítem en forma incorrecta (P(incorrecto)).

En el ensamblaje del instrumento final se utilizaron dos criterios: a) que las categorías definidas como rangos de frecuencia de aparición de las palabras en el corpus se ordenasen de forma ascendente, y b) que la probabilidad promedio de responder erróneamente los ítems dentro de cada categoría (promedio P(incorrecto), fuera menor al mínimo valor de P(incorrecto) del conjunto de ítems de la categoría de frecuencia superior. Por ejemplo, como se observa en la Tabla 4, la categoría de frecuencia con palabras cuya frecuencia es mayor a 100 veces en 100 libros tiene un promedio de P(incorrecto) de 0.1287, y la probabilidad mínima de las palabras de rango de frecuencias superior (40-100 veces en 100 libros) es 0.1333. Es decir, palabras como “grasa”, “cultivar” o “prohibido” tienen mayor probabilidad de ser respondidas en forma incorrecta que el promedio P(incorrecto) de la categoría inferior. Los ítems seleccionados utilizando estos criterios se muestran en la Tabla 4.

**Tabla 4. Ensamblaje prueba EVOC**

PALABRA	FRECUENCIA PROMEDIO EN 100 LIBROS	CATEGORÍA DE FRECUENCIA	P(INCORRECTO)	PROMEDIO P(INCORRECTO)
biblioteca	155		0.0792	
sano	115		0.0895	
viajar	189		0.0906	
estudiante	355		0.0906	
lanzar	224		0.0958	
rodeado	103	>100	0.1015	0.1287
poderoso	135		0.1136	
frente	1055		0.1521	
capítulo	269		0.1677	
querer	2161		0.2094	
inmenso	141		0.2255	
grasa	77		0.1333	
cultivar	87	40-100	0.1394	0.2049
prohibido	42		0.1566	

[CONTINÚA EN LA PÁGINA SIGUIENTE]

**Tabla 4. Ensamblaje prueba EVOC**

PALABRA	FRECUENCIA PROMEDIO EN 100 LIBROS	CATEGORÍA DE FRECUENCIA	P(INCORRECTO)	PROMEDIO P(INCORRECTO)
pata	46	40-100	0.1583	0.2049
abandonado	89		0.1646	
canta	53		0.1948	
serena	56		0.1997	
sostenido	73		0.1997	
tribu	47		0.2021	
condena	44		0.2031	
tomo	46		0.2781	
aislado	75		0.3146	
cubo	84		0.3188	
pesa	38	15-40	0.2094	0.4765
cojear	21		0.2385	
conmovido	36		0.3442	
consejero	28		0.3646	
intolerable	16		0.3683	
desembarcar	29		0.3833	
clamar	18		0.3941	
asistido	18		0.5125	
comprimir	21		0.5552	
perturbar	25		0.6187	
restauración	22	<15	0.6198	0.5734
perplejo	32		0.6292	
hazaña	34		0.6844	
contraído	29		0.7490	
gimnasia	9		0.4854	
recubierto	11		0.5052	
desolación	12		0.6187	
estandarte	12	0.6844		

*Criterio de corte*

Para determinar el criterio de punto de corte de la prueba, se estimaron las probabilidades de seguir avanzando en la prueba por azar, no sabiendo la respuesta del ítem. La estimación se realizó a través de cálculo de probabilidades en distribución binomial bajo las siguientes condiciones: a) tener, al menos, 1, 2, 3, 4, 5, 6, 7 y 8 ítems correctos al azar, en los últimos 8 ítems. b) tener, al menos, 1, 2, 3, 4, 5, 6 y 7 ítems correctos al azar, en los últimos 7 ítems, y c) obtener 1, 2, 3, 4, 5 y 6 ítems correctos al azar, en los últimos 6 ítems. La solución que permite disminuir la probabilidad de avanzar en la prueba por azar, desde cerca del 0.6 (obtener al menos 2 respuestas correctas dentro de 8 ítems) a debajo del 0.25, es la opción de detener la evaluación cuando existan 3 respuestas correctas en los últimos 7 ítems. Otra característica de esta solución es que se necesitan más respuestas incorrectas que correctas para detener la evaluación.

*Cálculo e interpretación de resultados*

Dado que se espera que, a medida que disminuye la frecuencia de aparición de las palabras en el corpus, éstas sean más difíciles, la interpretación de resultados considerará que, cuando el estudiante responda palabras más difíciles, tendrá mejores habilidades de vocabulario. En este sentido, si el evaluado alcanza un nivel de alta dificultad y frecuencia, pero no responde correctamente alguna pregunta de baja dificultad y frecuencia, las respuestas correctas de alta dificultad compensarán las respuestas correctas de baja dificultad (Cizek, 2012; Cizek y Bunch, 2007). Es decir, la habilidad de vocabulario no estaría dada por la suma total de respuestas correctas, sino más bien por aquel nivel en el que un estudiante obtuvo el corte en la prueba.

## DISCUSIÓN

El presente estudio exploró la entrega de una interpretación edumétrica de los resultados de la prueba EVOC, con el fin de que se utilicen en la toma de decisiones pedagógicas para incrementar el conocimiento léxico de los estudiantes. A través de los análisis realizados se estableció una interpretación basada en la frecuencia de aparición de cada palabra en un corpus de palabras provenientes de textos del currículo escolar chileno. El análisis factorial permitió eliminar ítems de baja carga al constructo principal, se establecieron índices de discriminación y dificultad por curso, y se ensamblaron los ítems en función de rangos de frecuencia de aparición de las palabras en el corpus escolar. Finalmente, al establecer nuevos criterios de

corte se redujo la probabilidad de responder correctamente por azar. Estos ajustes no solo constituyen mejoras psicométricas importantes que agregan validez al instrumento, sino que responden al objetivo de aprovechar mejor los resultados de los estudiantes en términos instruccionales. Las modificaciones realizadas en el presente estudio permitirán a los docentes interpretar los resultados de sus estudiantes en EVOC de manera que puedan determinar su capacidad de evocar el significado de las palabras en los textos leídos, y así puedan dimensionar mejor las necesidades léxicas específicas de cada uno de ellos.

Como señalan Kremmel y Schmitt (2016), se sabe poco acerca de cómo interpretar los resultados de pruebas de vocabulario, y por ende, del conocimiento léxico subyacente de los sujetos. En línea con este argumento, interesa saber si aquellos ítems de EVOC que los estudiantes responden en forma correcta son, efectivamente, palabras que serían capaces de comprender en un texto escrito, y si, tal como sostienen estos dos autores, rara vez se especifica el nivel y tipo de conocimiento que una prueba de vocabulario entrega (Bachman y Palmer, 2010).

En un estudio anterior (Orellana-García *et al.*, 2020) se explica cómo la primera versión de EVOC responde a sugerencias para mejorar la evaluación del vocabulario, que ya en 2007 surgían como necesarias para medir con mayor precisión el manejo léxico de los estudiantes, en particular en su relación con las habilidades de comprensión de texto (Pearson *et al.*, 2007). La selección de palabras a partir de corpus auténticos permite acercarse a una interpretación más concreta de lo que significa conocer las palabras evaluadas, y se podría deducir que, si los estudiantes identifican correctamente el significado de éstas en la prueba EVOC, también aplicaban su conocimiento de dicho significado a la hora de leerlas en el contexto de las obras incluidas en el corpus, que son también las lecturas a las que se enfrentarán en la enseñanza básica. Evidentemente, no se trata de una interpretación perfecta de los resultados en función de la habilidad de lectura, ya que, como sostienen Kremmel y Schmitt (2016), prácticamente no existe este tipo de estudios, dada la complejidad que supone evaluar el uso de la palabra para comprender un texto. Sin duda este tipo de información sería de gran utilidad para los docentes, pero creemos que la presente interpretación de los resultados de EVOC se aproxima a ese objetivo.

Un aspecto ampliamente debatido en la literatura es el uso de frecuencias de palabras versus el uso del ranking de dificultad. Como explicamos en el marco teórico, los corpus lingüísticos ordenan las palabras de mayor a menor frecuencia en un determinado campo textual, siendo esto atribuible a una mayor dificultad. En el caso de EVOC, la utilización del corpus lingüístico constituyó ya un avance significativo para contar con instrumentos de medición de dominio léxico más cercanos al ámbito de palabras que los escolares deben conocer para comprender textos, pero

en el presente estudio se depuró aún más el uso de la frecuencia al usar la frecuencia promedio de aparición de cada palabra en un rango de cantidad de libros. También se establecieron categorías de frecuencia en libros (de mayor a menor) y probabilidades de responderlas de forma incorrecta. Este tipo de indicadores, usados en forma conjunta para ensamblar los ítems, permitió evitar que ítems de menor frecuencia fuesen, efectivamente, más difíciles para los estudiantes.

Finalmente, el uso del formato de pregunta en test de vocabulario, la selección de distractores y la incidencia de estos dos aspectos en una mayor o menor proporción de respuestas correctas atribuibles al azar, continúan siendo temas importantes en la elaboración de este tipo de instrumentos. El uso de selección múltiple en los test de vocabulario es bastante común porque ofrece ventajas importantes, como la entrega de puntajes “objetivos”, una administración y corrección relativamente sencilla, y la posibilidad de evaluar a un mayor número de sujetos en menos tiempo (Gyllstad *et al.*, 2015). No obstante, este tipo de formato no está exento de debilidades. Una de ellas es la naturaleza de los distractores elegidos (Goodrich, 1977) y su impacto en el resultado total. En el caso de EVOC, el riesgo de la selección de distractores inadecuados se ha minimizado al establecer condiciones que delimitan claramente la selección de potenciales distractores, como son, no sólo la forma gramatical, sino también el ranking de frecuencia y la pertenencia a una misma categoría de frecuencia. Los resultados del estudio muestran que tomar en cuenta estos aspectos en función de los dos criterios estipulados (ordenación ascendente de los ítems en función de frecuencia promedio de aparición en el corpus, y probabilidad menor de responder erróneamente al mínimo valor de probabilidad en el conjunto de ítems de frecuencia superior) reduce la posibilidad de escoger un distractor que no funcione adecuadamente. Un segundo problema ampliamente discutido en investigaciones acerca de propiedades de instrumentos para medir vocabulario es la obtención de una respuesta correcta por adivinación o azar (Kremmel y Schmitt, 2016). En evaluaciones de vocabulario como el VLT (*Vocabulary Level Test*), de Nation y Belgar (2007), el porcentaje de respuesta correcta atribuida al azar es del 17%, lo que ocurre hasta un rango en el que el sujeto conoce más del 60% de los términos, punto en el que las respuestas correctas por azar comienzan a disminuir (Stewart y White, 2011). En el presente estudio se simuló respuestas al azar y, tal como se explica en la sección de resultados, al establecer criterios de corte de los test más rigurosos se minimizó el impacto de la respuesta por azar, disminuyendo considerablemente la probabilidad de que el estudiante avanzase en la prueba simplemente como resultado de adivinar. Desde un punto de vista edumétrico, estos ajustes permiten una interpretación de resultados que se acerca más al nivel real de dominio léxico de cada estudiante, al eliminar la respuesta correcta

por azar y contar con criterios de corte que bloquean la selección aleatoria de la respuesta correcta. Desde un punto de vista pedagógico, los resultados de EVOC proporcionan al docente una información valiosa respecto al manejo de palabras que cada estudiante tiene y cómo se relacionan con su frecuencia de aparición en textos que forman parte del currículum.

Una limitación de la versión actual del instrumento es que, si bien se reduce la probabilidad de que el evaluado avance en el test sin conocer realmente las palabras, aún es posible que adivine la opción correcta usando otras estrategias de resolución, como descartar distractores. Esto pretende ser evaluado empíricamente en aplicaciones posteriores, estimando el límite superior del error estándar de la desviación estándar del tiempo de respuesta de un set de ítems en aquellas personas con altos niveles de conocimiento de palabras. Este podría ser un límite a partir del cual se cree que el evaluado contestó a través de habilidades inferenciales.

Otra limitación es la cantidad no homogénea de ítems por banda de frecuencia en el corpus chileno. Algunos estudios sugieren que deben existir al menos 10 ítems, y hasta 30 ítems por banda de frecuencia (e.g. Gyllstad *et al.*, 2015). Dada la baja cantidad de ítems en la última banda de frecuencia, se hace necesario incorporar a la aplicación nuevas palabras con frecuencias menores, que serán piloteadas en próximas aplicaciones pero no consideradas para el cálculo de nivel de frecuencia del evaluado.

Finalmente, el hecho de que los indicadores de ajuste de los modelos factoriales no sean excelentes puede constituir una limitación que también podrá ser abordada pilotando nuevos ítems en aplicaciones posteriores. Esto permitirá agregar los ítems que se ajusten mejor a los datos.

Fecha de recepción del original: 31 de julio 2020

Fecha de aceptación de la versión definitiva: 10 de diciembre 2020

## REFERENCIAS

- Alber, K. (2016). Vocabulary Breadth, Vocabulary Depth, and Fluency -A theoretical framework and empirical evidence of lexical competence. *Zeitschrift Für Angewandte Linguistik*, 65(1). <https://doi.org/10.1515/zfal-2016-0019>.
- Anderson, R. C. y Freebody, P. (1981). Vocabulary knowledge. En J. T. Guthrie (Ed.), *Comprehension and Teaching: Research Reviews* (pp. 77-117). Newark: International Reading Association. <https://doi.org/10.4236/psych.2012.31007>.
- Anderson, R. C. y Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading. En P. D. Pearson, R. Barr, M. L. Kamil y P. Mosenthal (Eds.), *Handbook of Reading Research* (pp. 255-292). New York: Longman.

- Anthony, L. (2017). Corpus Linguistics and Vocabulary: A Commentary on Four Studies. *Vocabulary Learning and Instruction*, 6(2), 79-87. <https://doi.org/10.7820/vli.v06.2.Anthony>
- Bachman, L. F. y Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford: OUP.
- Barker, F., Salamoura, A. y Saville, N. (2015). Learner corpora and language testing. En S. Granger, G. Gilquin y F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Research* (pp. 511-534). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139649414.023>
- Beaujean, A. A. (2014). *Latent variable modeling using R: A step-by-step guide*. New York: Routledge.
- Biancarosa, G. y Griffiths, G. G. (2012). Technology tools to support reading in the digital age. *Future Child*, 22(2), 139-160. <https://doi.org/10.1353/foc.2012.0014>.
- Binder, K. S., Cote, N. G., Lee, C., Bessette, E. y Vu, H. (2017). Beyond breadth: The contributions of vocabulary depth to reading comprehension among skilled readers. *Journal of Research in Reading*, 40(3), 333-343. <https://doi.org/10.1111/1467-9817.12069>
- Bronfenbrenner, U. y Morris, P. A. (2006). The bioecological model of human development. En R. Lerner y W. Damon (Eds.), *Handbook of child psychology. Vol 1: Theoretical models of human development* (pp. 793-828). Hoboken: Wiley.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. New York: Guilford Publications.
- Chall, J. S. (1995). *Readability revisited: The new Dale-Chall readability formula*. Cambridge: Brookline Books.
- Cizek, G. J. (Ed.) (2012). *Setting performance standards: Foundations, methods, and innovations*. New York: Routledge.
- Cizek, G. J. y Bunch, M. B. (2007). *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests*. California: SAGE. <https://doi.org/10.4135/9781412985918>
- Daller, H., Milton, J. y Treffers-Daller, J. (2007). *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press.
- Duff, D., Tomblin, J. B. y Catts, H. (2015). The Influence of Reading on Vocabulary Growth: A Case for a Matthew Effect. *Journal of Speech, Language, and Hearing Research*, 58(3), 853-864. [https://doi.org/10.1044/2015\\_JSLHR-L-13-0310](https://doi.org/10.1044/2015_JSLHR-L-13-0310)
- Dunn, L. M. y Dunn, L. M. (1997). *Peabody picture vocabulary test*. Circle Pines: American Guidance Service.

- Echeverría, M. S., Herrera, M. O. y Segure, J. T. (1995). *TEVI-R Test de Vocabulario en Imágenes*. Concepción: Editorial Universidad de Concepción.
- Echeverría, M. S., Herrera, M. O. y Segure, J. T. (2002). *(TEVI-R) Test de Vocabulario en Imágenes. Manual de aplicación*. Concepción: Editorial Universidad de Concepción.
- Field, A. P., Miles, J. y Field, Z. (2012). *Discovering statistics using R*. California: SAGE.
- Fry, E. B. (1968). A readability formula that saves time. *Journal of Reading*, 11(7), 513-516. [https://www.jstor.org/stable/40013635?seq=1#metadata\\_info\\_tab\\_contents](https://www.jstor.org/stable/40013635?seq=1#metadata_info_tab_contents)
- Goodrich, H. C. (1977). Distractor efficiency in foreign language testing. *TESOL Quarterly*, 11(1), 69-78. <https://doi.org/10.2307/3585593>
- Gyllstad, H., Vilkaite, L. y Schmitt, N. (2015). Assessing vocabulary size through multiple-choice formats: issues with guessing and sampling rates. *Institut Voor Toegepaste Linguïstiek*, 166(2), 278-306. <https://doi.org/10.1075/itl.166.2.04gyl>
- Harrington, D. (2009). *Confirmatory factor analysis*. New York: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195339888.001.0001>
- Hiebert, E. H., Scott, J. A., Castaneda, R. y Spichtig, A. (2019). An Analysis of the Features of Words That Influence Vocabulary Difficulty. *Education Sciences*, 9(1), 8. <https://doi.org/10.3390/educsci9010008>
- Kamil, M. y Hiebert, E. (2005). Teaching and learning vocabulary: Perspectives and persistent issues. En E. H. Hiebert y M. L. Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 1-23). Mahwah: Lawrence Erlbaum.
- Kieffer, M. J. y Lesaux, N. (2008). The role of derivational morphology in the reading comprehension of Spanish-speaking English language learners. *Reading and Writing: An Interdisciplinary Journal*, 21(8), 783-804. <https://doi.org/10.1007/s11145-007-9092-8>
- Kremmel, B. y Schmitt, N. (2016). Interpreting Vocabulary Test Scores: What Do Various Item Formats Tell Us About Learners' Ability to Employ Words? *Language Assessment Quarterly*, 13(4), 377-392. <https://doi.org/10.1080/15434303.2016.1237516>
- Medellín, A. y Rodríguez, I. (2014). Propuesta metodológica para la evaluación de vocabulario académico a través de la lingüística de corpus. *RLA. Revista de lingüística teórica y aplicada*, 52(2), 41-63. <https://doi.org/10.4067/S0718-48832014000200003>
- Mendive, S., Mascareño Lara, M., Aldoney, D., Pérez, J. C. y Pezoa, J. P. (2020). Home Language and Literacy Environments and Early Literacy Trajectories

- of Low-Socioeconomic Status Chilean Children. *Child Development*, 1-21. <https://doi.org/10.1111/cdev.13382>
- Mineduc (2016). Centro de Estudios, División de Planificación y Presupuesto. Reporte Nacional de Chile: Revisión OCDE para mejorar el uso de recursos en las escuelas. Santiago de Chile: Ministerio de Educación. [https://centroestudios.mineduc.cl/wp-content/uploads/sites/100/2017/06/SCHOOL-final-impresion\\_OK.pdf](https://centroestudios.mineduc.cl/wp-content/uploads/sites/100/2017/06/SCHOOL-final-impresion_OK.pdf)
- Nagy, W. E., Anderson, R. C. y Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, 24, 237-270. <https://doi.org/10.2307/1162893>
- Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59-82. <http://dx.doi.org/10.3138/cmlr.63.1.59>
- Nation, P. y Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13.
- Orellana-García, P., Valenzuela-Hasenohr, M. F., Kung, M., Elmore, J. y Stenner, A. J. (2020). EVOC: un Instrumento para Evaluar Vocabulario en Español en Chile, *Magis, Revista Internacional de Investigación en Educación*, 13, 1-26. <https://doi.org/10.11144/Javeriana.m13.eiev>
- Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98, 554-566. <https://doi.org/10.1037/0022-0663.98.3.554>
- Park, K. (2014). Corpora and Language Assessment: The State of the Art. *Language Assessment Quarterly*, 11(1), 27-44. <https://doi.org/10.1080/15434303.2013.872647>.
- Pearson, P. D., Hiebert, E. H. y Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly*, 42, 282-296. <https://doi.org/10.1598/RRQ.42.2.4>
- Proctor, C. P., Silverman, R. D., Harring, J. R. y Montecillo, C. (2012). The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish-English bilingual children in elementary school. *Reading and Writing*, 25(7), 1635-1664. <https://doi.org/10.1007/s11145-011-9351-6>.
- Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading comprehension: an assessment perspective. *Language Learning*, 52(3), 513-536. <https://doi.org/10.1111/1467-9922.00193>
- Rasinski, T. y Rupley, W. (Eds.) (2019). *Vocabulary Development*. Special Issue in the Education Sciences Series. Basel: MDPI. <https://doi.org/10.3390/books978-3-03897-735-3>

- Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? En P. Bogaards y B. Laufer (Eds.), *Vocabulary in a Second Language: Selection, acquisition and testing* (pp. 209-227). Amsterdam: John Benjamins. <http://dx.doi.org/10.1075/llt.10.15rea>
- Read, J. (2019). Key issues in measuring vocabulary knowledge. En S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 545-560). Londres: Routledge.
- Richard, J. P. J. (2011). Does size matter? The relationship between vocabulary breadth and depth. *Sophia International Review*, 33, 107-120.
- Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech vocabulary development. *Child Development*, 83, 1762-1774. <https://doi.org/10.1111/j.1467-8624.2012.01805.x>.
- Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, 64, 913-951. <https://doi.org/10.1111/lang.12077>
- Schrank, F. A., McGrew, K. S., Ruef, M. L., Alvarado, C. G., Muñoz-Sandoval, A. F. y Woodcock, R. W. (2005). *Batería III Woodcock-Muñoz*. Rolling Meadows: Riverside Publishing.
- Simon, K. (2017). How I developed vocabulary tests using corpus-based word lists. *Prosperitas*, 4(4), 6-28. Budapest: BGE. <https://bit.ly/2HCP71T>.
- Stewart, J. y White, D. (2011). Estimating guessing effects on the vocabulary levels test for differing degrees of word knowledge. *TESOL Quarterly*, 45(2), 370-380. <https://doi.org/10.5054/tq.2011.254523>.
- Tannenbaum, K. R., Torgesen, J. K. y Wagner, R. K. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading*, 10, 381-98.
- Taylor, L. y Barker, F. (2008). Using Corpora for Language Assessment. En E. Shohamy y N. H. Hornberger (Eds.), *Encyclopedia of Language and Education. Volume 7: Language testing and assessment* (pp. 241-254). New York: Springer. [https://www.researchgate.net/deref/http%3A%2F%2Fdx.doi.org%2F10.1007%2F978-0-387-30424-3\\_179](https://www.researchgate.net/deref/http%3A%2F%2Fdx.doi.org%2F10.1007%2F978-0-387-30424-3_179)
- Thompson, P. y Sealy, A. (2007). Through children's eyes? Corpus evidence of the features of children's literature. *International Journal of Corpus Linguistics*, 12(1), 1-23. <https://doi.org/10.1075/ijcl.12.1.03tho>
- Torgesen, J. K., Wagner, R. K. y Rashotte, C. A. (1997). Prevention and remediation of severe reading disabilities: Keeping the end in mind. *Scientific Studies of Reading*, 1(3), 217-234. [https://doi.org/10.1207/s1532799xssr0103\\_3](https://doi.org/10.1207/s1532799xssr0103_3)
- Wright, T. S. y Cervetti, G. N. (2017). A Systematic Review of the Research on Vocabulary Instruction that Impacts Text Comprehension. *Reading Research Quarterly*, 52, 203-226. <https://doi.org/10.1002/rrq.163>

---

# RECENSIONES

