

The relation between reading comprehension and receptive vocabulary in Chilean students: an exploratory study

La relación entre comprensión lectora y vocabulario receptivo en estudiantes chilenos: un estudio exploratorio

José Pablo Pezoa

<http://orcid.org/0000-0002-4837-9289>

Universidad de los Andes (Chile)

Pelusa Orellana

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

Universidad de los Andes (Chile)

Received:
23/05/2020

Accepted:
14/01/2021

ISSN: 1885-446 X
ISSNe: 2254-9099

Keywords:
Reading comprehension;
reading processes; sight
vocabulary; vocabulary
development; reading abilities.

Palabras clave:
Comprensión lectora; procesos
de lectura; vocabulario
receptivo; desarrollo del
vocabulario; habilidades de
lectura.

Contact:
Jppezoa@uc.cl

Abstract

Although studies have described a non-causal relation between vocabulary and comprehension, the invariant effect of receptive vocabulary on comprehension has not been studied. The purpose of the current study is to determine whether the effect of receptive vocabulary on comprehension is invariant according to vocabulary level, sex and type of school among Chilean fifth-grade students. 553 Chilean fifth-grade students were assessed in reading comprehension and vocabulary. Using spline regressions, we observed that, when students get fewer correct responses on EVOC, a receptive vocabulary test, the effect of vocabulary on comprehension is null or smaller, than when students get higher vocabulary scores. These findings are useful when assessing the feasibility of interventions that may improve reading comprehension through vocabulary meaning.

Resumen

Si bien la bibliografía especializada ha descrito la existencia de un efecto posiblemente no causal entre vocabulario receptivo y comprensión lectora, y la posibilidad de un efecto acumulativo entre lectura de palabras y vocabulario, no se ha explorado si el efecto de vocabulario receptivo en comprensión lectora es invariante según los niveles de vocabulario receptivo, sexo y dependencia del colegio en la población chilena de quinto año básico. La muestra se conformó por 553 estudiantes balanceados en sexo y dependencia escolar, a los que se le aplicaron test de comprensión lectora y vocabulario. Los análisis realizados mediante regresiones *spline* muestran que, cuando los estudiantes obtienen puntajes en vocabulario receptivo cercanos a 30 ítems correctos en el instrumento EVOC, el efecto de vocabulario receptivo sobre comprensión lectora es nulo o mucho menor que cuando los estudiantes tienen mayores niveles de vocabulario receptivo. Estos hallazgos son un insumo a la hora de evaluar la factibilidad de intervenciones que busquen mejorar la comprensión a través de vocabulario.

This research has been funded by CORFO (Corporación de Fomento de la Producción) in the context of Project 18COTE-89646: Plataforma Tecnológica para la Evaluación de Habilidades de lectura.

Pezoa, J. P., & Orellana, P. (2021). The relation between reading comprehension and receptive vocabulary in Chilean students: an exploratory study. *Ocnos*, 20 (2), 7-20.
https://doi.org/10.18239/ocnos_2021.20.2.2407



Introduction and context

The process of learning to read and developing reading expertise is complex and does not end with decoding and fluency during the first years of schooling. Reading demands increase considerably at the beginning of the second primary school cycle (10- to 13-year-olds) and require students to have greater abilities for extracting, analysing, interpreting and evaluating written information (Stafura & Perfetti, 2015; Toste & Ciullo, 2017). Students progressively face a larger amount of informational texts, more sophisticated academic vocabulary, and much more complex reading tasks (Kendeou et al., 2008). Studies show that between 22% and 41% of students who have not been diagnosed with reading difficulties at the beginning of their schooling may show them in fourth or fifth grade (Leach et al., 2003), mainly due to the challenges involved in reading texts that are longer, have greater syntactic and semantic complexity, and encompass larger demands of conceptual knowledge.

National and international evaluations during the last decade show that Chilean students have low levels of comprehension throughout the different stages of their school life and that, although it is an issue across socioeconomic groups, the worst results are always observed in the most vulnerable contexts (Figueroa-Sepúlveda & Gallego-Ortega, 2018). On the other hand, the stagnation in reading proficiency as well as the gender and socioeconomic gaps are problems that have not been solved despite being prioritized from public policy. It is important to explore the relation between vocabulary and reading comprehension in fifth-grade students (10 years old), due to the fact that it is a crucial moment in elementary or primary education, before content area matter diversifies and texts increase in semantic complexity, all of which can affect the comprehension of disciplinary texts (Chall, 1983; Chall & Jacobs, 2003). For these reasons, fifth grade seems to be a level in which reading comprehension can still be increased by means of vocabulary interventions. This study seeks to categorize Chilean students according to their

reading comprehension level and explore the way these two constructs are related considering gender and type of school.

Literature Review

Reading performance of fifth-grade Chilean students

In Chile, reading comprehension is measured annually through standardized testing designed by the Sistema Nacional de Evaluación de Resultados de Aprendizaje (SIMCE) (National System for the Assessment of Learning Outcomes) in grades 4, 8 and 10. The SIMCE exam is applied by the Ministry of Education and measures the attainment of learning outcomes and curriculum coverage. In the area of reading, it assesses comprehension of different types of text and curriculum contents in the subject of Language and Communication. The items on reading comprehension, aligned with the current curricular framework, determine the students' reading comprehension proficiency, labeling them according to standards defined as initial, intermediate and advanced levels. According to the 2018 descriptors for fourth grade, the advanced level suggests the ability to relate and integrate explicit and implicit (or inferred) information, while the intermediate level consists in students being able to read and extract explicit information, making inferences from explicit text information. The initial level corresponds to those students who have not reached the intermediate level (Mineduc, 2008). Although reading comprehension levels have risen in the last decades, the increase has been rather poor. For example, for fourth grade in primary school, scores went up from 250 to 269 points between 2007 and 2017. However, when the data is interpreted in terms of attainment levels, we observe that, on average, students have stayed at the intermediate level.

Another assessment of reading skills in primary school is PIRLS (Progress in International Literacy Study). Chile participated in this international study for the first time in 2016, when the reading

skills of almost 4,300 fourth-grade students were evaluated. PIRLS measures reading comprehension skills with literary and informative texts at four levels of the reading processing: identifying information, making inferences, interpreting and integrating information, and analysing and evaluating the text's content (Agencia de Calidad de la Educación, 2017). The PIRLS scale has a mean score of 500 points, establishing 4 achievement benchmarks: low (under 400 points), intermediate (475 points or over), high (550 points or over), and advanced (625 points or over). Chilean students had an average of 494 points, but 26% performed at the lowest level and 13% below low. Only 25% of the assessed students were at the advanced level, which, similar to SIMCE, would show that there is an important amount of students whose reading comprehension is still deficient.

Reading comprehension gaps by gender and type of school

The performance of Chilean students at the middle of primary school shows a worrying stagnation in scores, and long-lasting socioeconomic and gender gaps that start in the first years of schooling, both in SIMCE and in PIRLS (Agencia de Calidad de la Educación, 2018). This is probably due to the impact of poverty in the development of oral language, particularly vocabulary, which affects children's reading performance and academic achievement in general (Hoff, 2003; Cain & Oakhill, 2006; Whitehurst, 1997), and relates to the system's inability to stop or revert these gaps. The studies that analyze the results of reading comprehension assessment systematically report statistically significant differences among students from public, subsidized and private schools, with better results for private and subsidized schools than those of public schools (Agencia de Calidad de la Educación, 2018). This is especially relevant in the Chilean context, since the type of school is associated to the families' payment ability and, therefore, their socioeconomic level. On the other hand, SIMCE and PIRLS results show a better performance in

women, compared to their male peers, in practically all grades year after year.

Relation between reading comprehension and vocabulary

The relation between vocabulary and reading comprehension is highly documented in the literature, with correlations that vary between .3 and .8 (Tannenbaum et al., 2006; Joshi, 2005). Although correlations might allude to possible influences, they do not define the nature of this relation as causal. Anderson and Freebody (1981) proposed three hypotheses to explain this relation: a) the instrumental hypothesis (the better the vocabulary, the better reading comprehension level), b) the knowledge hypothesis (a non-causal correlation that also relates to conceptual knowledge), and c) the aptitude hypothesis, which proposes a non-causal correlation with verbal aptitude (Cain & Oakhill, 2011; Oullette, 2006; Nagy, 2007, among others). In the case of the instrumental hypothesis, for instance, a reader who does not know the meaning of 2% to 5% of the words in a text will struggle to understand it (Carver, 1994; Oslund et al., 2016). The effect of vocabulary on comprehension is not only found at the level of individual words, but at that of larger units, such as sentences and paragraphs (Elbro & Buch-Iversen, 2013). Knowing the meaning of a word allows the reader to establish connections between ideas that are in separate sentences and integrate knowledge as they read, making inference generation easier (Oslund et al., 2016). Therefore, vocabulary has an indirect impact on comprehension through its influence at a sentence comprehension level, which, in turn, promotes the textual cohesion with which inferences are made along the text. Anderson and Freebody (1981) propose an additional indirect impact of vocabulary in their second hypothesis by establishing a relation between the knowledge of a word's meaning and a network of conceptual knowledge that the reader retrieves in order to understand a text by means of connections and inferences. If a particular word is part of the reader's

vocabulary, it facilitates the activation of associated concepts that allow to relate ideas in the text, and, for those students with a lower cultural capital, lack of vocabulary would account for their lower ability to formulate inferences (Keenan et al., 2014). Sternberg and Powell (1983) also link vocabulary to reading comprehension through inferences, which Nagy (2007) defines as metalinguistic awareness. The theoretical framework developed by Perfetti (1999) also proposes a causal relation underlying the vocabulary and comprehension relation, and which is backed by the view of comprehension as a meaning-building process through the integration of the information of the text and previous knowledge, including knowledge of the meaning of words (Kintsch, 1998; van den Broek, 2010). On the other hand, there are some who propose that the relation between these two concepts is founded upon the common processes both share (Nation, 2009; Cain et al., 2004). An increase in vocabulary requires the ability to map or relate meaning and phonological form, while the rise in reading comprehension would require a similar relation among units of semantic, phonological and orthographic representation (Perfetti, 2007). However, the relation between comprehension and vocabulary is not just found at word-level but also in larger units (Nation, 2008).

Ultimately, the relation between vocabulary and reading comprehension is mediated by an integration process between the text and previous knowledge, such as word meaning, allowing for the inference of meanings of larger units like sentences (Perfetti & Stafura, 2014). In the context of reading, in order to infer meaning from units longer than a word – as the case of the sentence – the reader needs at least the ability of evoking the meaning of a relatively high percentage of the words in the text (Carver, 1994; Oslund et al., 2016), which allows for the activation of associated concepts that contribute to the construction of meaning (Keenan et al., 2014). Hence, although it is known that the better the level of knowledge of words, the greater the comprehension (Tannenbaum et al., 2006; Joshi, 2005), it is not known whether this relation is different

according to the student's word knowledge levels. This is relevant for cost-benefit analyses in interventions that seek to promote reading comprehension through the increase of knowledge of words. For example, if we know that from a given level of word knowledge onwards there is not a sufficiently big effect on reading comprehension, it would not be beneficial to raise the costs of those programs. On the other hand, if we know that an increase in word knowledge has an effect on – or explains – reading comprehension, only from a specific point in the level of command of words, it is justifiable to intervene in order to overcome the threshold from which word knowledge begins to explain reading comprehension, privileging its benefits.

The current study

This study seeks to characterize the current situation of fifth-year students according to sex and type of school, and explore the way reading comprehension and receptive vocabulary are related according to sex and school type. These results are an important input when evaluating the suitability of interventions looking to promote reading comprehension through an increase in vocabulary at the end of elementary or primary education.

Methods

Participants

The participants are 553 fifth-grade students (10 years old) from 9 different schools. 139 students come from public schools (25.14%), 275 from subsidised schools (49.7%) and 139 from private schools (25.1%). Of all these students, 227 are male (50.1%) and 276 are female (49.9%). The average in EVOC, a receptive vocabulary test, was 51.8 points of a maximum of 64 points.

Instruments

In order to identify the students' reading level, we used the Dialect test. This test was originally

elaborated in Spanish and its validity is supported by empirical and theoretical evidence (Orellana & Melo, 2015). Dialect has 40 reading comprehension questions in Spanish, where students read brief texts of increasing complexity, to then answer a multiple-choice question. The questions require students to infer the right answer from the text. Given that the test was made using the framework proposed by MetaMetrics, the interpretation of the scores is done in terms of Lexile measures, where the first items of the test have texts with lower Lexile text measures, becoming more complex as the test progresses (Scholastic, 2014).

Furthermore, we applied the EVOC instrument, measuring receptive vocabulary; that is, the ability of relating a word with a representative image of the corresponding object (Orellana-García et al., 2020). The instrument consists in having students listen to the target word and are then showing them four images on the screen. They have to select the image corresponding to the meaning of the word they heard. The stimuli and the distractors were obtained from a linguistic corpus built from all of the written material (textbooks, complementary readings, documents and others) to which Chilean students are exposed in primary school. These 64 words are organized by difficulty, with increasing complexity. The instrument follows the parameters of similar assessments (e.g., Dunn & Dunn, 2007; Echeverría et al., 2001) such as stopping the evaluation after the test-taker makes six mistakes in eight consecutive items. EVOC has been validated for the Chilean population (Orellana-García et al., 2020).

Procedure

For the collection of data, we gathered the information of 553 fifth-grade students, with the appropriate consents, in the Metropolitan region. The participants were assessed on their vocabulary (EVOC) and reading comprehension (Dialect) towards the end of the school year. Each student answered the Dialect and EVOC tests using a digital platform under the supervision of the teachers at

each school. The application of the test took about 35 minutes on average.

Analysis

In order to explore the performance of fifth-grade Chilean students, and characterise them in terms of school type, sex and EVOC scores, we built descriptors as well as frequency and student percentage tables by performance category in the Dialect test. To evaluate whether the frequency of students in each Dialect performance level depends on sex or type of school, we did a chi-squared test (Agresti, 2018). When this analysis drew significant results, we carried out a log-linear analysis. This analysis estimates gross residuals, from which the adjusted residual is estimated, statistic that shows how random or not the joint frequency is. It allows to know in which cells of the contingency table there is a greater or smaller observed frequency than that expected at random; or concretely: to know whether a given level of reading comprehension is associated to sex or type of school (Bakeman & Robinson, 2013).

Although it is known that the greater the knowledge of words, the better the reading comprehension, it is not known if this relation is completely linear, a linear spline, or quadratic. This study explores the shape of said relationship, for which the explained variance of linear regression, quadratic regression and spline regression models is compared. A spline regression can be understood as a model “that consists of a continuous explanatory variable defined over specified segments of the domain of that variable, and a dependent variable that is a continuous function of that explanatory variable over all segments, but with different slopes in each of the segments” (Marsh & Cormier, 2002). Using this type of analysis allows to find two regressions with a single breakpoint (or knot) through minimising square residuals (Lunt, 2000), which, in turn, allows to know if there is any inflection point in the linear relation between vocabulary and reading comprehension. The model that explains the greatest amount of variation will be chosen as the best.

Results

Sample

The sample consists of 553 students assessed through Dialect, EVOG and characterization questions, which have scores in Dialect. Out of the total of the students considered in the analysis, there are no missing values in the variables of sex or type of school. There are 65 cases with missing data in the EVOG variable that equal 11.8% of the sample. In order to evaluate whether the missing values are random, we carried out a t-test for independent samples and variances to assess if the average levels of reading comprehension are different between the groups with or without missing values. The test shows that the averages of reading comprehension in Lexiles between the groups with or without missing values are significantly different ($t = 8.09, p < .000$). To know if this difference in scores is related to belonging to a type of school, we did a chi-squared analysis entering the variable of missing data in EVOG and type of school. The analysis shows that having missing values depends on the type of school ($\chi^2 = 77, p < .000$). There is a higher percentage of missing values in public schools and a lower percentage of missing values in private ones, regarding the total number of students in each type of school (table 1). This reveals that the analyses that involve EVOG are more generalizable for private and subsidised schools than for public schools.

Table 1
 Missing data percentage by type of school

Type of school	Not missing %	Missing %
Public	67.63	32.37
Subsidised	93.82	6.18
Private	97.84	2.16

Descriptors

As table 2 shows, fifth-grade students had an average level of reading comprehension of 640L, which, according to the bands determined by Scholastic for the population in the United States (Scholastic, 2014), places them at the basic level (between 620L and 829L for fifth grade). Moreover, students got 52 points on average in the EVOG test, which places them between the 47 and 49 percentiles of the validation sample (Orellana-García, et al., 2020). The sample consists of almost 50% males and 50% females, and most students attended subsidised schools (table 2).

Table 2
 Lexile, EVOG and characterization descriptors

Variable	Frequency	Average	SD	Min	Max
Lexiles	533	640	295	27	1291
EVOG	488	52	7	19	62
	Frequency	Average			
Sex					
Female	276	49.9			
Type of school					
Public	139	25.1			
Subsidised	275	49.7			
Private	139	25.1			

Reading comprehension performance

According to the proficiency bands developed by Scholastic (2014) for the American population, the majority of Chilean students are below the basic level (45%). Only 25% of students are at the proficient or advanced level (table 3).

Table 3
 Fifth grade reading comprehension level

Level	Frequency	Average
Below basic	249	45.03
Basic	169	30.56
Proficient	64	11.57
Advanced	71	12.84
Total	553	100

Characterisation of reading comprehension levels

The t-test analysis for independent samples of reading comprehension between men and women shows that there are no significant differences between males and females in terms of Lexiles ($t = -1.8, p = .069$). On the other hand, the omnibus test shows that there are, in fact, significant differences among the Lexile averages by type of school ($F = 91.36, p < .000$). We did a Tukey's post-hoc test between the averages in reading comprehension among students that attend different types of school. The analysis shows that the difference in the reading comprehension averages between public and subsidised schools is not significant ($p = .063$); however, the differences in reading comprehension between public-private schools and subsidised-private schools are significant ($p < .000$) (table 4).

In order to deepen our understanding of the characterization of students according to their performance in comprehension, contingency tables were created

including the levels of reading comprehension and sex, and between reading comprehension and type of school. We also tested whether the proficiency level in reading comprehension depends on sex and/or type of school by means of a chi-squared test. The analysis shows that the proficiency level observed in reading comprehension in fifth grade does not depend on whether the student is male or female ($\chi^2 = 4.47, p = .214$); however, the proficiency level in reading comprehension in fifth grade does depend on the type of school the student attends ($\chi^2 = 148.5, p < .000$).

To know which levels of reading comprehension are related to a particular type of school, a log-linear analysis (Bakeman & Robinson, 2013) was carried out. The analysis shows that the frequency observed of students that achieve an advanced level in reading comprehension at public schools is lower than the frequency expected. On the other hand, the frequency observed in students from subsidized schools who achieve a below basic level is higher than expected, and lower than the expected frequency for students that achieve an advanced level of reading comprehension. The frequency observed in students from private schools is lower at the below basic and basic levels and higher than expected at the proficient and advanced levels (table 5).

Shape of the relation between comprehension and vocabulary

With the purpose of exploring the shape of the relation between reading comprehension and vocabulary, we compared the explanatory variance of different models that explain the relation between

Table 4
 Tukey's Post Hoc contrasts of Lexiles by type of school

Comparison		Contrast	SE	t	p	CI	
						min	max
Subsidised	Public	-60.1	26.6	-2.3	0.063	-122.7	2.5
Private	Public	294.6	30.7	9.6	0	222.4	366.8
Private	Subsidised	354.7	26.6	13.3	0	292.1	417.3

Table 5
 Percentage of students by proficiency level and sex

Level	Sex		Type of school		
	Male	Female	Public	Subsidised	Private
	Frequency (%)	Frequency (%)	Frequency (%)	Frequency (%)	Frequency (%)
Below basic	134	115	71	155⁺	23[·]
	54	46	29	62	9
Basic	74	95	48	91	30[·]
	44	56	28	54	18
Proficient	31	33	11	16[·]	37⁺
	48	52	17	25	58
Advanced	38	33	9 [·]	13 [·]	49 ⁺
	54	46	13	18	69

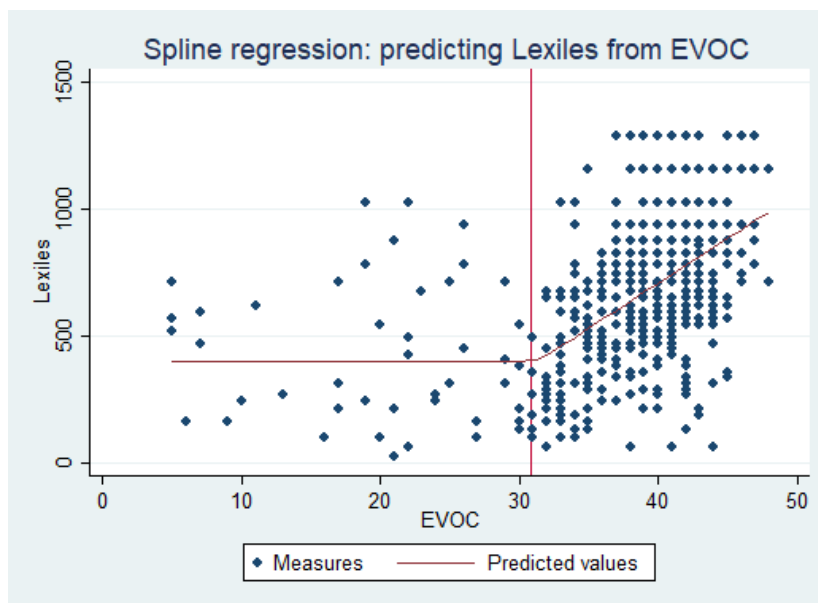
Table 6
 Model comparison between reading comprehension and vocabulary

Models	Predictors	Coefficient	SE	t	p	CI		Adjusted R2
						min	max	
Linear								0.188
	Intercept	-13.4	64.8	-0.2	0.8	-140.7	113.9	
	EVOC	18.0	1.7	10.7	0.0	14.7	21.3	
Quadratic								0.261
	Intercept	682.4	117.7	5.8	0	451.0	913.7	
	EVOC	-33.9	7.7	-4.4	0	-49.0	-18.9	
	EVOC ²	0.9	0.1	7.0	0	0.6	1.1	
Spline								0.275
	Intercept	502.5	102.8	4.9	0	300.6	704.5	
	Knot	30.8	1.7	17.9	0	27.4	34.2	
	Slope 1	-4.1	4.7	-0.9	0.388	-13.3	5.2	
	Slope 2	36.1	3.1	11.5	0	30.0	42.3	

reading comprehension and receptive vocabulary (table 6). As shown in table 6, the reading comprehension linear function regressed in receptive vocabulary explains 19% of the variance of the reading comprehension scores. By adding a quadratic coefficient to the model, the explanatory

variance of the model rises to 26%. The spline model is the one that explains a greater variance (28%). This model reveals that when students have fewer than 31 right answers in EVOC (45 points), the EVOC test does not significantly explain reading comprehension measured through Dialect. With a

Figure 1
Spline regression of receptive vocabulary predicting reading comprehension



score higher than 45, represented by item 31 of the test, the relation between vocabulary and reading comprehension is significant and linear (figure 1).

When the spline regression is modeled by sex, we observe that the males who on average exceed 31 correct items (45 points in EVOc), achieve 42L on average for each right answer in vocabulary, whereas the females who exceed the same number of correct items in vocabulary, only benefit from 27L for each right answer in vocabulary. In other words, males increase their reading comprehension about 35% more than females assessed for each right answer in vocabulary.

In terms of school type, public schools begin to gain in reading comprehension from lower levels of receptive vocabulary than subsidized and private schools (27 items versus 30 and 33, respectively). On the other hand, the schools that benefit the most from a rise in receptive vocabulary are subsidized and public ones with 30L and 27L for each right answer in EVOc, from their knots. Private schools increase 26L per right answer in EVOc from their knot onwards (table 7).

Discussion and conclusions

The results of this exploratory study show low levels of reading comprehension performance among Chilean assessed fifth-grade students. According to these results, almost half of the students is at below basic level and the largest proportion of students with a low reading proficiency is concentrated in schools with fewer resources. Just as in measurements such as SIMCE and PIRLS, there is a significant gap between school types, not so much in terms of gender, which differs from other evaluations (Agencia de Calidad de la Educación, 2017; 2018).

Another important finding of the study is the way in which receptive vocabulary and reading comprehension are related. The spline model shows that receptive vocabulary has an effect on reading comprehension (possibly non-causal) when the level of receptive vocabulary surpasses 45 points (31 correct items). This results in that, for each additional point scored in receptive vocabulary, students increase 36L in their reading comprehension (18L more than that explained by a linear regression). This would allow for the assumption

Table 7
 Model comparison between reading comprehension and vocabulary

Models	Predictors	n	Coefficient	SE	t	p	CI		Adjusted R2
							min	max	
Men Spline		244							0.34
	Intercept		539.4	125.4	4.3	0.000	292.2	786.5	
	Knot		30.1	2.0	15.3	0.000	26.2	34.0	
	Slope 1		-8.5	6.1	-1.4	0.163	-20.6	3.5	
	Slope 2		42.2	4.5	9.3	0.000	33.3	51.1	
Women Spline		248							0.2
	Intercept		500.3	178.5	2.8	0.005	148.8	851.9	
	Knot		32.1	3.0	10.6	0.000	26.1	38.1	
	Slope 1		-0.6	7.2	-0.1	0.931	-14.7	13.5	
	Slope 2		29.8	4.6	6.5	0.000	20.8	38.7	
Public school Spline		94							0.21
	Intercept		751.8	231.4	3.3	0.002	292.0	1211.5	
	Knot		27.0	3.8	7.1	0.000	19.4	34.5	
	Slope 1		-14.8	11.6	-1.3	0.207	-37.8	8.3	
	Slope 2		26.9	6.1	4.4	0.000	14.8	38.9	
Subsidised school Spline		258							0.28
	Intercept		344.3	122.3	2.8	0.005	103.5	585.0	
	Knot		30.2	2.8	10.8	0.000	24.7	35.7	
	Slope 1		-0.9	5.9	-0.2	0.874	-12.6	10.7	
	Slope 2		30.4	3.6	8.4	0.000	23.3	37.6	
Private school Spline		136							0.14
	Intercept		687.2	200.0	3.4	0.001	291.5	1082.9	
	Knot		32.9	5.7	5.7	0.000	21.5	44.2	
	Slope 1		0.2	9.0	0.0	0.987	-17.7	18.0	
	Slope 2		25.8	6.5	4.0	0.000	12.9	38.7	

that there are students who achieve a level of word meaning knowledge that helps them make better inferences in continuous texts, thus improving their performance in comprehension tasks that require inferring. These results are consistent with those observed by Stafura & Perfetti (2014), Oslund et al. (2016), and Cromley & Azevedo (2007) regarding the existence of an indirect relation between vocabulary and reading comprehension and the fact that this is due to the contribution of knowledge

of word meaning to the formulation of inferences during the comprehension process in longer texts. This study's findings further extend the insight provided by the aforementioned studies, suggesting the existence of a cumulative effect on the relation between vocabulary and reading comprehension; i.e. a threshold from which accumulated vocabulary knowledge drastically improves reading comprehension levels. We also observe that males benefit more than females in reading comprehension when

considering the point in which receptive vocabulary begins to have an effect on reading as the starting point of the linear estimate. This shows that, in the case of male students, pedagogical interventions for increasing their receptive vocabulary level could have a greater effect on reading comprehension than females. Similar evidence is observed in the study by Price-Mohr & Price (2016), who suggest that male children profit more than females from the use of hybrid methods (e.g. synthetic and holistic) in the learning of reading, and that they learn more quickly with the use of authentic books with vocabulary that includes more complex words than those they are able to decode. Likewise, studies have demonstrated that male children better optimize the use of working memory, for instance, when understanding subtitled narrations (Linebarger, 2001), and that they prefer visual-spatial learning (Wang & Carr, 2014), which would be important to consider from the instruction perspective. Future research will be able to examine the next processes through which these findings can be further explained. Moreover, public-school students have a lower vocabulary threshold from which they benefit from an increase in word knowledge than those in subsidized and private schools. This is important due to the smaller increments at the level of vocabulary knowledge, as it may have greater impact on reading comprehension, if and when this increment goes over the 27-point threshold on the EVOC test.

The findings also bring forth implications for public policy design. Given that the slope in the relation between vocabulary and reading comprehension is greater from a specific level of achievement in receptive vocabulary, it would seem more beneficial to design policies that tend to increase this vocabulary to levels over the threshold of 31 correct items in the receptive vocabulary measured with EVOC, which could have an effect on reading comprehension. In contrast, not purposefully taking EVOC scores beyond the threshold could have no effects on reading comprehension and widen gaps of a potential Matthew effect in vocabulary (Duff et al., 2015; Stanovich, 1986). Moreover, analyses suggest that public-school students could show improvement with a lower level of vocabulary than

students in subsidized and private schools, and that, in a context with limited resources or schools different types of school, but with the same level of vocabulary, it may be more suitable to focus resources on interventions in public schools. Future studies may further examine the processes that produce a cumulative effect of vocabulary on reading comprehension.

Limitations

This study was centered on the assessment of comprehension skills and vocabulary in fifth-grade students (10-11 years old), which does not allow for a wider view of the relation between vocabulary and reading comprehension throughout primary education. On the other hand, although the sample is of a reasonable size, there is missing data regarding public schools. This constitutes a limitation in the generalization of findings, since it distorts the representativeness of the population.

References

- Agencia de Calidad de la Educación (2017). *Informe Nacional de Resultados PIRLS 2016*. Agencia de Calidad de la Educación (Santiago de Chile). http://archivos.agenciaeducacion.cl/PIRLS_V02_27MAR.pdf.
- Agencia de Calidad de la Educación (2018). *Resultados Estudio Nacional de Lectura Segundo Básico*. Agencia de Calidad de la Educación (Santiago de Chile). http://archivos.agenciaeducacion.cl/IRE_LECTURA_2018_2BASICA_WEB_ALTA_11_JUL.pdf.
- Agresti, A. (2018). *An introduction to categorical data analysis*. John Wiley & Sons.
- Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77-117). International Reading Association.
- Bakeman, R., & Robinson, B. F. (2013). *Understanding log-linear analysis with ILOG: An interactive approach*. Psychology Press. <https://doi.org/10.4324/9780203763216>.
- Biemiller, A., & Slomin, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology*, 93(3), 498-520. <https://doi.org/10.1037/0022-0663.93.3.498>

- Bowyer-Crane, C., & Snowling, M. J. (2005). Assessing children's inference generation: What do tests of reading comprehension measure? *British Journal of Educational Psychology*, 75(2), 189-201. <https://doi.org/10.1348/000709904X22674>.
- 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>.
- Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual Differences in the Inference of Word Meanings From Context: The Influence of Reading Comprehension, Vocabulary Knowledge, and Memory Capacity. *Journal of Educational Psychology*, 96(4), 671-681. <https://doi.org/10.1037/0022-0663.96.4.671>.
- Cain, K., & Oakhill, J. (2006). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology*, 76(4), 683-696. <https://doi.org/10.1348/000709905X67610>.
- Carver, R. P. (1994). Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: Implications for instruction. *Journal of Literacy Research*, 26(4), 413-437. <https://doi.org/10.1080/10862969409547861>.
- Chall, J. (1983). *Stages of Reading Development*. McGraw Hill.
- Chall, J., & Jacobs, V. (2003). The classic study on poor children's fourth-grade slump. *American Educator*, 27(1), 14-15.
- Chile. Mineduc, Unidad de Curriculum y Evaluación (2008). *Niveles de logro Cuarto Básico para lectura*. http://archivos.agenciaeducacion.cl/biblioteca_digital_historica/orientacion/2008/nl4b_lectura_2008.pdf.
- Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, 99(2), 311-325. <https://doi.org/10.1037/0022-0663.99.2.311>.
- Duff, D., Tomblin, J.B., & 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.
- Dunn, L. M., & Dunn, D. M. (2007) *PPVT-4: Peabody picture vocabulary test*. Pearson Assessments. <https://doi.org/10.1037/t15144-000>.
- Echeverría, M., Herrera, M.O., & Segure, M. (2001). *TEVI-R Test de Vocabulario en Imágenes*. Editorial Universidad de Concepción.
- Elbro, C., & Buch-Iversen, I. (2013). Activation of background knowledge for inference making: Effects on reading comprehension. *Scientific Studies of Reading*, 17(6), 435-452 <https://doi.org/10.1080/10888438.2013.774005>.
- Figueroa-Sepúlveda, S., & Gallego-Ortega, J. L. (2018). Vocabulario y comprensión lectora en escolares chilenos de primer ciclo básico. *Ocnos*, 17(1), 32-42. https://doi.org/10.18239/ocnos_2018.17.1.1521.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Brookes.
- Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, 74(5), 1368-1378. <https://doi.org/10.1111/1467-8624.00612>.
- Joshi, M. R. (2005) Vocabulary: A Critical Component of Comprehension, *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 21(3), 209-219. <https://doi.org/10.1080/10573560590949278>
- Keenan, J., Hua, A., Hulstander, J., Christopher, M, & Olsen, R. (2014, July). *How do sources of individual differences in reading comprehension vary with age and assessment?* Poster presented at the annual meeting of the Society for the Scientific Study of Reading, Santa Fe, New Mexico. <https://www.triplesr.org/how-do-sources-individual-differences-reading-comprehension-vary-age-and-assessment>.
- Kendeou, P., Bohn-Gettler, C., White, M., & Van de Broek, P. (2008). Children's inference generation across different media. *Journal of Research in Reading*, 31(3), 259-272. <https://doi.org/10.1111/j.1467-9817.2008.00370.x>.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.
- Leach, J.M., Scarborough, H.S., & Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology*, 95(2), 211-224. <https://doi.org/10.1037/0022-0663.95.2.211>.
- Linebarger, D. L. (2001). Learning to read from television: The effects of using captions and narration. *Journal of Educational Psychology*, 93(2), 288-298. <https://doi.org/10.1037/0022-0663.93.2.288>.

- Lunt, M. (2000). *Piecewise linear ("Hockey-Stick") regression*. <https://personalpages.manchester.ac.uk/staff/mark.lunt/loghockey.html>.
- Marsh, L. C., & Cormier, D. R. (2002). *Quantitative Applications in the Social Sciences: Spline regression models*. SAGE Publications. <https://doi.org/10.4135/9781412985901>.
- Nagy, WE. (2007). Metalinguistic awareness and the vocabulary-comprehension connection. In R. Wagner, A. Muse, K. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 53-77). Guilford Press.
- Nation, K. (2009). Reading comprehension and vocabulary: What's the connection? In R. K. Wagner, C. Schatschneider, & C. Phythian-Sence (Eds.), *Beyond decoding. The behavioral and biological foundations of reading comprehension* (pp. 176-193). Guilford Press.
- Nation, I. S. P. (2009). *Teaching ESL/EFL reading and writing*. Routledge. <https://doi.org/10.4324/9780203891643>.
- Orellana, P., & Melo, C. (2015). Dialect: Integrating technology and reading assessment to diagnose Spanish reading difficulties. *Journal of Literacy and Technology*, 16(1), 38-66.
- 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. <https://doi.org/10.11144/Javeriana.m13.eiev>.
- Ouellette G. (2006). What's meaning got to do with it: the role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98(3), 554-566. <https://doi.org/10.1037/0022-0663.98.3.554>.
- Oslund, E. Clemens, N. H., Simmons, D. C., Smith, S. L., & Simmons, L. E. (2016). How vocabulary knowledge of middle-school students from low socioeconomic backgrounds influences comprehension processes and outcomes. *Learning and Individual Differences*, 45, 159-165. <https://doi.org/10.1016/j.lindif.2015.11.013>.
- Perfetti, C. A. (1999). Comprehending written language: A blueprint of the reader. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 167-208). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198507932.003.0006>.
- Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension, *Scientific Studies of Reading*, 11(4), 357-383. <https://doi.org/10.1080/10888430701530730>.
- Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, 18(1), 22-37. <https://doi.org/10.1080/10888438.2013.827687>.
- Price-Mohr, R., & Price, C. (2017). Gender Differences in Early Reading Strategies: A Comparison of Synthetic Phonics Only with a Mixed Approach to Teaching Reading to 4-5 Year-Old Children. *Early Childhood Education Journal*, 45, 613-620. <https://doi.org/10.1007/s10643-016-0813-y>.
- Psyridou, M., Eklund, K., Poikkeus, A-M., & Torppa, M. (2018). Reading outcomes of children with delayed vocabulary: A follow-up from age 2-16. *Research in Developmental Disabilities*, 74, 114-124. <https://doi.org/10.1016/j.ridd.2018.05.004>.
- Scholastic (2014). *SRI College and Career Technical Guide. Using a valid and reliable assessment for College and Career Readiness in grades K-12*. Scholastic https://www.hmhco.com/products/assessment-solutions/assets/pdfs/sri/SRI_TechGuide.pdf.
- Stafura, J. Z., & Perfetti, C. A. (2014). Word-to-text integration: Message level and lexical level influences in ERPs. *Neuropsychologia*, 64, 41-53. <https://doi.org/10.1016/j.neuropsychologia.2014.09.012>.
- Stafura, J. Z., & Perfetti, C. A. (2015). Comprehending implicit meanings in text without making inferences. In E. J. O'Brien, A. E. Cook, & R. F. Lorch (Eds.), *Inferences during reading* (pp. 16-33). Cambridge University Press.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21(4), 360-407. <https://doi.org/10.1598/RRQ.21.4.1>.
- Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. Guilford.
- Sternberg, R. J., & Powell, J. S. (1983). Comprehending verbal comprehension. *American Psychologist*, 38(8), 878-893. <https://doi.org/10.1037/0003-066X.38.8.878>.
- 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. https://doi.org/10.1207/s1532799xssr1004_3.
- Toste, J. R., & Ciullo, S. (2017). Reading and Writing Instruction in the Upper Elementary Grades. *Intervention in School and Clinic*, 52(5), 259-261. <https://doi.org/10.1177/1053451216676835>.

- van den Broek, P. (2010). Using texts in science education: Cognitive processes and knowledge representation. *Science*, 328(5977), 453-456. <https://doi.org/10.1126/science.1182594>.
- Van Vreckem, C., Desoete, A., & Van Keer, H. (2011). Poor comprehensive readers: What do we know about their profile? *Procedia Social and Behavioral Sciences*, 15, 229-234. <https://doi.org/10.1016/j.sbspro.2011.03.078>.
- Wang, L., & Carr, M. (2014). Working memory and strategy use contribute to gender differences in spatial ability. *Educational Psychologist*, 49(4), 261-282. <https://doi.org/10.1080/00461520.2014.960568>.
- Whitehurst, G. J. (1997). Language processing in context: Language learning in children reared in poverty. In L. B. Adamson, & M. A. Romskey (Eds.), *Research on communication and language disorders: Contribution to theories of language development* (pp. 233-266). Brookes.