



## Vocabulary depth and its contribution to text quality in the early years of primary school (*Profundidad de vocabulario y su contribución a la calidad textual en los primeros años de Educación Primaria*)

M. Dolores Alonso-Cortés-Fradejas, Mercedes López-Aguado & M. Teresa Llamazares-Prieto

To cite this article: M. Dolores Alonso-Cortés-Fradejas, Mercedes López-Aguado & M. Teresa Llamazares-Prieto (2021) Vocabulary depth and its contribution to text quality in the early years of primary school (*Profundidad de vocabulario y su contribución a la calidad textual en los primeros años de Educación Primaria*), Journal for the Study of Education and Development, 44:1, 82-116, DOI: [10.1080/02103702.2020.1848087](https://doi.org/10.1080/02103702.2020.1848087)

To link to this article: <https://doi.org/10.1080/02103702.2020.1848087>

 View supplementary material 

 Published online: 06 Jan 2021.

 Submit your article to this journal 




 Article views: 157

 View related articles 

 View Crossmark data 



## Vocabulary depth and its contribution to text quality in the early years of primary school (*Profundidad de vocabulario y su contribución a la calidad textual en los primeros años de Educación Primaria*)

M. Dolores Alonso-Cortés-Fradejas , Mercedes López-Aguado   
and M. Teresa Llamazares-Prieto 

Universidad de León

### ABSTRACT

The purpose of this study is to analyse how vocabulary depth develops in the early years of primary school and to ascertain the effect it has on the quality of the texts produced by the students. A total of 332 Spanish students from the first to fourth grades participated by doing a synonyms/antonyms task to evaluate their vocabulary depth, along with several tasks in which they wrote descriptive texts to evaluate text quality. Vocabulary depth was analysed using a four-dimensional system created ad hoc: change in lexeme, fitting the meaning requested, preservation of syntactic category and degree of conventionality. The text quality indicators were productivity, subordination rate and text organization. The results revealed that both vocabulary depth and the quality of the texts produced improved significantly by grade. The study also found that this type of lexical knowledge contributes significantly to the quality of the texts evaluated. Finally, this article discusses the possible teaching implications of these findings.

### RESUMEN

El propósito de este estudio es analizar cómo evoluciona la profundidad de vocabulario durante los primeros cursos de Educación Primaria y comprobar qué efecto tiene en la calidad de los textos producidos por el alumnado. Participaron 332 estudiantes españoles, de 1º a 4º de Educación Primaria, que llevaron a cabo una tarea de sinónimos/antónimos, para evaluar la profundidad del vocabulario, y varias tareas de escritura de textos descriptivos para evaluar la calidad del texto. La profundidad de vocabulario se analizó con un sistema de cuatro dimensiones creado ad hoc: cambio de lexema, ajuste al sentido del significado solicitado, preservación de la categoría sintáctica y grado de convencionalidad. Como indicadores de la calidad textual se eligieron la productividad, el índice de subordinación y la estructura textual. Los resultados revelaron que tanto la profundidad del vocabulario como la calidad de los textos producidos mejoraron significativamente con el nivel escolar. Se encontró también que este tipo de conocimiento léxico contribuye significativamente a la calidad textual

### ARTICLE HISTORY

Received 29 April 2019  
Accepted 25 January 2020

### KEYWORDS

vocabulary; text quality; text organization; synonyms; antonyms

### PALABRAS CLAVE


vocabulario; calidad textual; estructura textual; sinónimos; antónimos

**CONTACT** M. Dolores Alonso-Cortés-Fradejas  [mdalof@unileon.es](mailto:mdalof@unileon.es)  Universidad de León, León, España.

English version: pp. 82–97 / *Versión en español*: pp. 98–114

References / *Referencias*: pp. 114–116

Translation from Spanish / *Traducción del español*: Mary Black

 Supplemental data for this article can be accessed [here](#) / Se puede acceder a los datos suplementarios para este artículo [aquí](#).

de los textos evaluados. Se comentan las posibles implicaciones didácticas de estos hallazgos.

One of the main objectives of primary school is for students to be able to effectively write texts with different characteristics depending on their communicative purpose; consequently, a series of learning outcomes which should be reached at each grade have been established. However, an analysis of the Spanish curriculum, for example, reveals that it does not state the expected progression in acquiring the learning outcomes, nor does it provide information on the factors which contribute to text quality at these grades.

It has been demonstrated that the vocabulary knowledge of students in the first grade explains more than 30% of the variation in their reading comprehension level in the last grade of primary school (Cunningham & Stanovich, 1997), but does this type of knowledge have the same explanatory power with regard to written expression? The different models that have been put forth to explain the text construction process concur that two types of knowledge come into play: content and language (lexical, syntactic, rhetorical) (Tolchinsky, 2015). To compose a text on a certain topic, writers have to be capable of choosing pertinent information; that is, they have to master the content. Yet they also have to know how to choose the most appropriate terms to express this content, and they have to know, for example, what other meanings these terms may have depending on the context and how they are associated with each other and with others. However, even though lexical knowledge is unquestionably one of the fundamental resources when writing a text, the majority of studies have tended to focus on its relationship with reading comprehension (Laufer & Ravenhorst-Kalovski, 2010; Ouellette & Beers, 2010; Sénéchal et al., 2006), and its potential contribution to the quality of written texts has yet to be explored. On the other hand, it has been found that while the capacity to decode is related to the number of words a person knows (*vocabulary breadth*), the capacity to access the meaning of texts depends on the vocabulary depth acquired, that is, on their skill of manipulating the meaning of known words (Ouellette, 2006). What effect can we expect these two dimensions of lexical knowledge to have on the production of written texts?

The purpose of this study is to analyse how *vocabulary depth* develops in primary school students and to ascertain whether it has an effect on the quality of the texts they produce. This will enable us to better determine the features of text quality in this population and therefore establish realistic learning outcomes which will, in turn, improve teaching processes.

### Text quality: indicators and skills involved

Describing the texts produced by schoolchildren is a very complicated undertaking; however, analysing how their quality develops is even more complex. In Spain, it has been observed that linguistic features (such as syntactic complexity, cohesion and the use of vocabulary) explain holistic perceptions of quality better than textual features (Salas et al., 2016). However, it is necessary to go beyond the holistic evaluation by external judges and to seek an analytical evaluation of the characteristic features at the different linguistic levels (word, sentence, text) (Verheyden et al., 2010).

The numerous indicators that have been used in different studies on the quality of the text produced by primary school students include linguistic features like lexical

richness (Malvern et al., 2004) and text organization features (Breetvelt et al., 1994; Tolchinsky, 2019).

These text features result from the combination of the subject-writer's different skills, including local linguistic expression related to lexical choice or sentence syntax, along with more global discursive skills such as the capacity to organize a text properly (Berman & Nir-Sagiv, 2009). Both have been used to try to explain text quality, but the results do not match (for a meta-analysis see Kent & Wanzek, 2016) and only account for one-third of the differences found (Salas & Tolchinsky, 2017). For example, Abbott et al. (2010) point out that spelling mastery is the variable that best explains longitudinal differences in text composition in the first seven years of primary school in English. However, Wagner et al. (2011) found that when explaining the developmental differences between English-speaking students in the first and fourth grades, handwriting fluency was the subjects' skill with the most explanatory power, while mastery of spelling was less explanatory. With regard to Spanish, in a study conducted on the same sample as this study, Tolchinsky (2019) found that the capacity to produce self-sustained oral discourse, to handle the relationships of meaning among words and good working memory explained the differences in the descriptive texts among the groups, while changes in the development of text organization were explained by a strong mastery of spelling.

### Vocabulary knowledge

During the lexical development process, the strategies that children use to learn the meanings of words and relate them to each other change as their internal representation of the language grows and is reorganized (Alexander Pan & Uccelli, 2010). Children continue to learn new words after their early years, but they also begin to learn other meanings of the words they already know and to make connections between them. That is, as their vocabulary expands it also becomes deeper. Consequently, beyond studying how many words learners know and how their vocabulary increases as they go through school, it is also essential to ascertain what relationships they have established between words and how these relationships change over time, given that this is also an important factor in their academic progress (Verhallen & Schoonen, 1998).

Standardized tests of expressive and/or receptive vocabulary are usually used to evaluate vocabulary breadth, or the number of lexical entries that a learner is capable of using (Castillo & Tolchinsky, 2018). However, evaluating vocabulary depth, or a learner's knowledge of the multiple meanings of words and their associations (Alexander Pan & Ucelli, 2010), is much more complicated since it is a dimension of lexical development that could be viewed as the mastery of one or several of the following types of knowledge and skills which overlap and are interrelated (Ordóñez et al., 2002; Schmitt, 2014): knowledge of the meaning of polysemic words, synonyms and antonyms, words from the same word family, semantic collocations (the words or types of words that can and/or should appear with a given word), lexical fluency (the capacity to access their own lexicon quickly and accurately) and lexical organization (the ability to recognize and provide word associations).

In consequence, current studies on the acquisition of lexical depth are scarce and do not always concur. For example, the results of the study by Kieffer and Lesaux (2012)

with sixth-grade speakers of English as a first and second language suggest that the distinction between vocabulary breadth and depth does not accurately reflect individual differences in any of the learners. However, Schwartz and Katzir (2012) detected that after one year of schooling, a group of bilingual seven-to-eight-year-old second-generation Russian-Hebrew speakers with a favourable socio-educational context showed improvement in both vocabulary breadth and depth. Therefore, it is essential to continue researching this aspect not only to determine its importance with regard to vocabulary breadth but also to ascertain how it changes over time, that is, from a longitudinal perspective.

### Relationship between vocabulary and text quality

When they plan, write or revise a text, writers have to choose the most appropriate words, according to their meaning, from among those that they know. Therefore, different studies have examined a series of aspects related to vocabulary breadth as indicators of text quality. One of them, lexical diversity, that is, the number of different words over the total number of words, is the feature analysed the most, and it has been found to be the lexical measure that best explains developmental differences between English-speaking students in the second and fourth grades (Olinghouse & Leaird, 2009). However, it is also known that when trying to explain variations in quality according to the type of text, other aspects of vocabulary breadth come into play: for persuasive texts, the number of thematic words — that is, the number of words related to the theme being written about; for informative texts, the number of thematic words and lexical maturity — that is, the number of words used that are not among the most frequent in each language (Olinghouse & Wilson, 2013).

On the other hand, during the text-writing process, writers have to be able to establish a series of semantic, morphological and syntactic relationships among the words in their lexicon (Castillo & Tolchinsky, 2018). However, these factors, which have to do with vocabulary depth, do not tend to be included among the quality indicators of written expression. Only a few studies, such as the one conducted by Gómez-Vera et al. (2016) with fourth-grade Chilean students, have demonstrated that the features of vocabulary depth explained the variation in quality better than lexical diversity and sophistication.

The contribution that vocabulary knowledge makes to the quality of written texts has not been explored much either, even though the results once again suggest that its effect is different according to whether vocabulary breadth or depth is being examined. In two studies conducted with Turkish-speaking students from the first to fifth grades, Babayiğit and Stainthorp (2010, 2011) concluded that along with working memory, vocabulary, measured in terms of breadth, was the only reliable longitudinal predictor of the quality of the text content, although they did not find that this variable contributed to the organizational quality of those texts. However, two studies conducted on texts produced by primary school students in Catalan (Castillo & Tolchinsky, 2018) and Spanish (Tolchinsky, 2019), respectively, concluded that vocabulary depth, measured as knowledge of synonyms and antonyms, improved with educational level and contributed positively to text organization. These latter results seem to indicate that vocabulary depth is a variable that has more explanatory potential when examined as

a predictive factor than as an indicator. Knowledge of synonyms and antonyms is a manifestation of the relationships that the subject-writer establishes paradigmatically, that is, between words that may or may not appear in the same place, unlike other dimensions such as thematic vocabulary or lexical diversity, which account for the relationships of meaning that the subject-writer establishes among the words that appear within the same text.

### **This study**

Although not exhaustive, the literature review examining the quality of texts produced by primary school students shows a lack of studies which outline how text quality develops in Spanish throughout the grades and that determine with which of the subject skills it is the most closely associated. Another conclusion of the literature review is that within these skills, vocabulary depth, measured as knowledge of synonyms and antonyms, has strong explanatory power as a predictive factor, although more evidence is needed. Therefore, two objectives were set for this study:

- (1) To explore how the level of vocabulary depth and the quality of the text produced develops in the early years of primary school, the latter through productivity (or number of words), syntactical complexity and text organization indicators.
- (2) To ascertain whether the level of vocabulary depth has an effect on the quality of the texts produced.

The hypotheses guiding the research were:

- (1) The level of vocabulary depth, specified in the use of synonyms and antonyms, improves with grade level.
- (2) As the grade level goes up, the texts will be better organized, have more words and be syntactically more complex.
- (3) Students with a greater vocabulary depth, specified in the use of synonyms and antonyms, will write longer, better organized and syntactically more complex texts.

## **Method**

### **Design**

To achieve these objectives, we chose a descriptive developmental methodology combining a longitudinal and cross-sectional study. Given that this study did not involve intervention, the teaching-learning process was not considered a variable.

### **Participants**

An intentional sample of 332 monolingual Spanish-speaking students from seven schools in four cities in Spain (Almería, Cádiz, Ciudad Real and Valencia) took part

in the study, and information was collected in two rounds one year apart. A total of 164 students (45.1% girls) in the first grade and 168 students (48.8% girls) in the third grade participated in the first year ( $M = 8.71$ ). In the second round, these students were in the second and fourth grades, respectively.

### **Tasks and instruments**

The tasks to collect information on vocabulary depth and text quality were carried out in two identical rounds.

### **Vocabulary depth**

Given that to evaluate text quality the participants were going to be asked to produce different descriptive texts and this would purportedly entail using many adjectives correctly, it was decided that vocabulary depth would be evaluated with a task that consisted of providing a synonym or antonym for those sets of adjectives. The adjectives were chosen from the CesCa corpus using a frequency criterion and were translated into Spanish, as this is a corpus of school texts in Catalan (Llaurado et al., 2012). They all belonged to the same semantic field (personality traits).

Furthermore, we chose an oral production task instead of multiple choice since vocabulary depth should be viewed as a continuum that encompasses everything from some degree of familiarity with the word to total ownership of the different meanings and the associated capacity to use them appropriately in different contexts (Kieffer & Lesaux, 2012). On the other hand, it was understood that for the participants, this task implied the same lexical skill involved in text production, that is, choosing words not from a set of alternatives provided but from their own mental lexis.

The researchers presented each of the participants with the same inventory of 28 words, one by one, and they had to orally produce a synonym. Then, using the same procedure, they presented another inventory of 28 words, the same one for each participant, and they had to orally produce an antonym. The instructions were as follows: ‘Now I’m going to say a series of words and you’ll have to tell me a synonym for each of them, that is, a word that has the same meaning; Now I’m going to say a series of words and you’ll have to tell me an antonym for each of them, that is, a word that means the opposite’. To ensure that they understood the task, examples with three synonyms and three antonyms were provided. The task was carried out individually in a room away from their usual classroom, there were no time restrictions and at no time did they have to read or write.

### **Text quality**

The text-production tasks were conducted after the evaluation of vocabulary depth. In the first round, each participant wrote three descriptive texts: ‘describe a good teacher’, ‘describe the house of your dreams’ and ‘describe a dangerous situation’. The texts that they were asked to write in the second round were ‘describe your best friend’ and ‘describe your ideal vacation spot’. The students were asked to write descriptive texts for two reasons: it is the only kind of text that appears in all primary school curricula in the

different regions in Spain as an evaluable learning standard, and even though descriptive texts have unique organizational features, they can also be integrated into other types of texts (such as narratives) that students at these grades are supposed to learn. The participants worked at the same time in their usual classroom, but the texts were written individually, as the outcome not of a learning task but of an evaluation task, each with a maximum of 20 minutes and with no more than a two-week interval between them. The participants composed a total of 492 texts in the first round and 336 in the second one, and all texts were digitalized.

### **Coding and analysis procedure**

Prior to the analysis, the responses obtained in the vocabulary depth task and the text produced by the participants were coded.

For the first coding, the research team created a system of eight categories based on a combination of four dimensions, which reproduce some of the operations involved in eliciting synonyms and antonyms; they were also applied in Castillo and Tolchinsky (2018) and Tolchinsky (2019):

#### **Change in lexeme**

The participant provides a different word than the word given.

#### **Fitting the meaning requested**

The participant provides a word with an equivalent meaning when they are asked for a synonym or a word with the opposite meaning when they are asked for an antonym.

#### **Preservation of the part of speech**

The participant provides an adjective.

#### **Conventionality**

The participant provides a word that appears as a synonym or antonym in four dictionaries taken as references.

In turn, in order to obtain a performance score on this task, and therefore on vocabulary depth, a scoring system from one to five was established, according to how many and what dimensions were manifested in each response: change in lexeme, 1 point; fitting the meaning requested, 2 points; preservation of the part of speech, 1 point; and conventionality, 1 point.

Therefore, when they were asked for a synonym or antonym, if the participants responded with a different adjective with an equivalent or opposite meaning, respectively, and that word appears as a synonym or antonym in the dictionaries consulted, this response was included within category 8 (see additional material): for example, *contento* (contented) as a synonym of *alegre* (cheerful) or *callado* (quiet) as an antonym of *charlatán* (talkative). The responses in this category were scored with a 5 because they reflected maximum knowledge of the semantic operations required to establish relations of antonymy and synonymy (fitting the meaning requested), lexical knowledge (change in lexeme and conventionality) and grammatical knowledge (preservation of part of speech).



However, when they were asked for a synonym or antonym, if the participants offered a different adjective with an equivalent or opposite meaning, respectively, but that word did not appear as a synonym or antonym in the dictionaries consulted, such as *feliz* (happy) as a synonym of *alegre* (cheerful) or *silencioso* (silent) as an antonym of *charlatán* (talkative), the responses were included in category 7. The responses in this category were scored with four points because they reveal solid semantic and grammatical knowledge, although they did not show awareness of the conventional lexical associations contained in the dictionaries. Likewise, given the frequency with which they appeared, it was deemed that responses that fulfilled the features mentioned above as well as the following should be included in this category: (a) they were the result of a metaphor or comparison — such as *fideo* (noodle)<sup>1</sup> as a synonym of *delgado* (thin) or *enano* (dwarf) as an antonym of *alto* (tall); (b) they could be considered valid in a certain register or variety — such as *machote* (beefcake) as a synonym of *fuerte* (strong) or ‘*cagao*’ (chicken) as an antonym of *valiente* (brave); or (c) they could be considered valid among children, such as *bueno* (good) as a synonym of *listo* (clever) (since adults often talk about a ‘good, clever child’) or *enfadado* (angry) as an antonym of *guapo* (handsome) (since adults in Spain often comment that ‘you’re not very handsome when you’re angry’).

The answers in which two or more dimensions were missing were included in the remaining categories.

When they responded to a request for a synonym or antonym with a different word that reflected the meaning requested, but it was not an adjective and consequently did not appear in the dictionaries as such, the responses were included in category 6, such as when they said *miedo* (fear) as a synonym of *cobarde* (coward) or *pesimismo* (pessimism) as an antonym of *optimista* (optimistic). The responses in this category were given a score of 4 because, just like the responses in category 7, they reflected lexical and semantic knowledge, albeit less grammatical knowledge.

When they responded with a different word that reflected the meaning requested, but that word could not be considered Spanish as it showed certain morphological changes, such as *\*hablatón* (\*speakative) for *charlatán* (talkative) or *\*calentero* (\*warmy) as an antonym of *friolero* (chilly), the participants’ words were included in category 5 and scored with a 3.

When asking them for a synonym and antonym, if the participants provided responses that were coherent with the meaning requested but they were not adjectives but phrases or sentences, these responses were included in category 4 and were scored with a 3: for example, *no lo sabe hacer* (they don’t know how to do it) as a synonym of *inseguro* (insecure) or *echarse para atrás* (back down) as an antonym of *decidido* (determined). In contrast, when they provided phrases or sentences but they included the word they were being asked about, for example, *llora mucho* (cries a lot) as a synonym of *llorón* (cry-baby) or *que no es aburrido* (not boring) as an antonym of *aburrido* (boring), the responses were included in category 3 and scored with a 2 since the only one of the four dimensions they showed was fitting the meaning requested.

The participants’ responses with an equivalent or opposite meaning, depending on the task, which did not entail a change in lexeme but did show some morphological variation, correct or not, were included in category 2 and scored with a 1, since they were considered to reflect more grammatical knowledge because they preserved the part

of speech: for example, *rabioso* (really angry) as a synonym of *rabioso* (angry) or *\*intriste* (\*unsad) as an antonym of *triste*. Finally, category 1 included all the responses considered wrong since they did not fit the meaning requested: for example, *enfadada* (angry) as a synonym of *alegre* (joyful) or *malón* (thug) as a synonym of *malo* (bad).

Thus, the maximum score possible was 280 points (28 synonyms x 5 points maximum score for each; 28 antonyms x 5 points maximum score for each). Before coding the entire sample, an internal validation process was carried out by assessing inter-judge agreement. The degree of agreement among the three expert judges was higher than 90% in all cases.

Furthermore, the texts produced by the participants were coded according to the following aspects:

### **Productivity**

Number of words written, that is, the number of chains of letters bounded by blank spaces on either side.

### **Subordination rate**

Number of subordinate clauses, except complete clauses, over the total number of clauses.

### **Text organization**

Presence or absence (1/0) of the functional components of descriptive texts: introductory, descriptive, supporting and concluding clauses. The score was given to each text and then the arithmetic mean was calculated for each round. A scale of 0 to 8 was used, as described below (Table 1):

To answer the first research question (objective 1), that is, to assess the developmental changes in the different variables, we analysed the effect of grade at school. Simple analysis of variance (ANOVA) tests were performed when the analysis was between different subjects and *t*-tests of related samples when the differences being analysed referred to the different rounds with the same subjects.

To answer the second research question (objective 2), that is, to explain the relationships among the variables chosen, we chose the model described in Figure 1. We hypothesized the existence of a latent variable called *text quality* which would be influenced by the observed variable *vocabulary depth*, and which would be explained by three observed variables: *number of words*, *subordination rate* and *text organization*.

We checked the organization proposed using a structural equation model, and we analysed whether this explanatory structure is equivalent throughout the different grades by a multigroup analysis. To avoid the effects of learning the test, only the responses from the first round were used (first and third grades).

The comparison between groups was established progressively following the instructions of Byrne (2004), and the fit indexes were analysed for each of the models proposed. In the first model, unconstrained factorial invariance was estimated, assuming differences in the groups comprising the sample. The second model constrained measurement weights, equalizing them in all groups, which meant considering that the weights of each item by factor are equivalent. The third model added a constraint on the effect of the structural weights; in other words, the effect of vocabulary depth on text

**Table 1.** System of evaluating and scoring text organization.

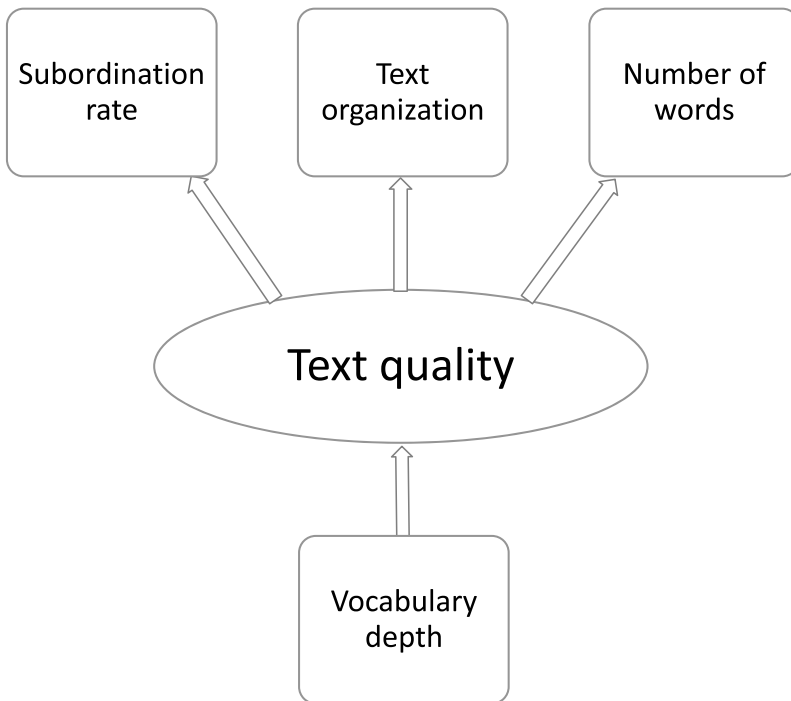
Score	Introduction	Descriptive	Support	Conclusion
8	1	1	1	1
7	0	1	1	1
6	1	1	1	0
5	0	1	1	0
4	0	1	0	1
3	1	1	0	0
2	0	1	0	0
1*	1	0	0	0
	0	0	1	0
	0	0	0	1
0	0	0	0	0

Note:

0 Absent

1 Present

1\*absence of descriptive clauses and presence of one of the other three types of clauses



**Figure 1.** Proposed model to explain text quality and its relationship to vocabulary depth.

quality was equalized. The fourth model added constraints on the structural covariances, assuming equality in the variances and covariances within the groups. Finally, the fifth model adds a constraint in the measurement residuals, assuming the equivalency of errors.

To determine the goodness of fit, the following indices and criteria were used: Chi square ( $\chi^2$ ) higher than .05; Chi square ratio between degrees of freedom ( $\chi^2/df$ ) lower than 3; root mean square error of approximation (RMSEA) lower than .05;

and the comparative fit index (*CFI*), Tucker-Lewis index (*TLI*) and normed fit index (*NFI*) higher than .95.

## Results

### *Development of vocabulary depth*

As shown in Table 2, the responses on the semantic task varied as the grades rose. Thus, there was a steady decrease in the responses in categories 1, 2, 3 and 4, which means that the respondent did not know how to apply or only partially applied the different operations entailed in eliciting a synonym or antonym: choosing a different word, also an adjective, with an equivalent or opposite meaning, respectively.

On the other hand, the correct responses steadily increased, that is, those in categories 7 and 8, which reveal mastery of these operations as well as — in the responses in category 8 — knowledge of the conventional relationships of meaning, that is, those contained in dictionaries.

Categories 5 and 6 both showed residual values in all four grades.

These developmental differences in vocabulary depth (among the four groups) are significant for both the synonyms (Chi-square = 1,027.87; *df* = 21; sig. < .01) and the antonyms (Chi-square = 1,229.83; *df* = 21; sig. < .01).

### *Development of the indicators of text quality*

Table 3 describes the differences in the mean scores on text organization, number of words and subordination rate for each of the rounds of one of the texts. As illustrated, the effect of the students' grade was significant (< .01), in favour of the highest grade in all the variables and all the texts, in both the first and second rounds. The results reveal

**Table 2.** Response percentage for each of the eight categories of synonyms and antonyms by grade.

		Grade			
		First cohort		Second cohort	
		1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>	4 <sup>th</sup>
Categories Synonyms	1	10.6%	8.8%	4.7%	5.1%
	2	11.3%	10.1%	7.6%	4.2%
	3	41.7%	34.6%	35.3%	23.4%
	4	2.0%	1.0%	1.3%	0.3%
	5	0.3%	0.2%	0.7%	0.5%
	6	1.4%	1.8%	1.6%	1.5%
	7	10.9%	14.9%	15.3%	21.0%
	8	21.8%	28.6%	33.7%	43.9%
Categories Antonyms	1	7.1%	6.1%	3.3%	2.7%
	2	3.0%	4.4%	3.8%	4.5%
	3	44.7%	28.3%	28.0%	14.4%
	4	0.6%	0.3%	0.5%	0.1%
	5	1.5%	1.8%	3.0%	2.6%
	6	0.8%	0.8%	0.8%	0.6%
	7	14.7%	20.6%	21.3%	26.7%
	8	27.4%	37.9%	39.3%	48.2%

**Table 3.** Effect of grade on text organization, number of words, subordination rate and vocabulary depth. ANOVA.

		Text organization					
		First round			Second round		
		Text 1	Text 2	Text 3	Text 1	Text 2	Text 3
1 <sup>st</sup>		3.35	2.46	3.67	4.44	3.98	
3 <sup>rd</sup>		4.32	3.80	4.83	5.16	4.69	
Total		3.85	3.14	4.27	4.80	4.34	
		$F = 18.90; df = 321; p < .01$			$F = 7.86; df = 293; p < .01$		
		$F = 42.07; df = 311; p < .01$			$F = 6.82; df = 268; p < .01$		
		$F = 24.90; df = 312; p < .01$			$F = 35.69; df = 313; p < .01$		
		First round			Second round		
		Text 1	Text 2	Text 3	Text 1	Text 2	Text 3
1 <sup>st</sup>		26.48	32.09	26.60	60.27	41.50	
3 <sup>rd</sup>		67.32	72.76	59.24	89.71	61.56	
Total		47.34	52.93	42.63	74.99	51.62	
		$F = 162.19; df = 322; p < .01$			$F = 24.53; df = 318; p < .01$		
		$F = 111.06; df = 321; p < .01$			$F = 32.5; p < .01$		
		$F = 147.99; df = 325; p < .01$			$F = 9.41; df = 316; p < .01$		
		Subordination rate					
		First round			Second round		
		Text 1	Text 2	Text 3	Text 1	Text 2	Text 3
1 <sup>st</sup>		29.91	19.61	30.42	30.50	33.95	
3 <sup>rd</sup>		39.12	29.67	38.18	39.12	35.12	
Total		34.66	24.65	34.36	34.99	34.54	
		$F = 10.81; df = 319; p < .01$			$F = 0.14; df = 316; p > .05$		
		$F = 13.30; df = 306; p < .01$			$F = 6.93; df = 324; p < .01$		

that in the participating sample, all the indicators chosen to assess the quality of text production steadily rose at higher grades.

To assess the developmental changes, we compared the scores of each participant in two consecutive grades. The results of the *t*-test for related samples indicated that statistically significant developmental gains were achieved in text organization, both between first and second grade [ $n = 156$ ,  $SD = -1.16$  (3.13–4.29),  $CI$  95% (–1.51, –0.82),  $t = -6.70$ ,  $df = 155$ ,  $sig < .01$ ] and between the third and fourth grades [ $n = 162$ ,  $SD = -0.57$  (4.35–4.92),  $CI$  95% (–0.91, –0.24),  $t = -3.36$ ,  $df = 161$ ,  $sig < .01$ ], and in number of words, both between the first and second grades [ $n = 159$ ,  $SD = -22.76$  (27.74–50.50),  $CI$  95% (–26.88, –18.63),  $t = -10.90$ ,  $df = 158$ ,  $sig < .01$ ] and between the third and fourth grades [ $n = 162$ ,  $SD = -9.38$  (66.06–75.44),  $CI$  95% (–14.54, –4.22),  $t = -3.59$ ,  $df = 162$ ,  $sig < .01$ ].

These results indicate that there were individual changes in the quality of the written output, which were reflected in better text organization and more words. However, even though there were differences in the means of the grades in the two rounds for the subordination variable, and the scores increased at higher grades, the analysis of individual differences in this indicator shows that they are not significant ( $t = 0.80$ ,  $sig. = .42$ ).

### Effect of vocabulary depth on text quality

We suggest that the relationships among the variables described above reflect the structure illustrated in [Figure 1](#), which hypothesizes:

- The existence of the latent variable *text quality*.
- This latent variable would be explained by three observed variables: *number of words*, *subordination rate* and *text organization*.
- The observed variable vocabulary depth has a direct effect on the latent variable text quality and an indirect effect on number of words, subordination rate and text organization.
- There are differences in the model of relationships between text quality and vocabulary depth depending on the grade.

To check these hypotheses, we conducted a multigroup analysis. The results of the fit data ([Table 4](#)) indicated that the unconstrained model was a good fit for all the indicators analysed. The significance associated with the Chi square was higher than .05, the quotient between this indicator and the degrees of freedom was lower than 3, the *RMSEA* was lower than .05 and the *CFI*, *TLI* and *NFI* were higher than .95, which indicates that the data collected fit the explanatory model proposed successfully.

The results support the existence of the latent variable text quality and confirmed that it is explained by the score on text organization and, to a lesser (albeit also significant) extent, by the number of words and subordination rate. They also confirmed the effect of vocabulary depth on text quality, with a direct standardized effect higher than .04 (.45 in the first grade; .43 in the third), which is considered an adequate value for considering one variable's effect on another. In other words, the results indicated that greater vocabulary depth determines a higher quality of texts produced.

**Table 4.** Comparison of model fit. Multigroup analysis.

Model	$\chi^2$ (sig.)	df	$\chi^2/df$	RMSEA	CFI	TLI	NFI
Unconstrained	5.93 (.204)	4	1.48	.038	.988	.939	.967
Metric invariance	18.06 (.006)	6	3.01	.078	.924	.746	.899
Factorial invariance	18.07 (.012)	7	2.58	.069	.930	.800	.899
Structural covariance	19.50 (.012)	8	2.43	.066	.927	.818	.891
Residual values	97.69 (.000)	11	8.88	.155	.453	.005	.452

We also confirmed that the two grades show differences in the model of relationships between text quality and vocabulary depth. All the goodness of fit indexes declined as the measurement weights, structural effects, structural covariances and measurement residuals equalized. None of these alternatives reached adequate values for a proper fit.

Table 5 describes the regression coefficients and the standardized effects on the two grades together to facilitate their comparison. The results show that at higher grades, the direct effect of vocabulary depth on text quality declines slightly, as does the indirect effect of this variable on the three quality indicators, especially on the number of words.

Conversely, as the grades get higher the model of relationships between text quality and vocabulary depth seems to increase based on the three variables proposed.

## Discussion

The purpose of this study was twofold. On the one hand, we sought to trace the development of a specific kind of linguistic knowledge, lexical knowledge, measured in terms of vocabulary depth, and the development of text quality, measured at word level (number of words), sentence level (syntactic complexity) and text level (text organization), in the early grades of primary school. To do so, we analysed the results by groups of first, second, third and fourth graders and the differences in the transition from first to second and from third to fourth grades. Likewise, we sought to determine whether vocabulary depth is a variable that influences the quality of the texts produced by students, and whether there are differences in this effect in relation to grade. To do so, we performed a multigroup structural covariance analysis with the students from the first and third grades to control for the effect of task repetition.

**Table 5.** Standardized regression coefficients and total direct and indirect standardized effects. First and third grades.

		Standardized effects			
		1 <sup>st</sup>		3 <sup>rd</sup>	
		VD	Quality	VD	Quality
Text quality	← Vocab. depth	.45		.43	
Subordination rate	← Text quality	.24*	.53	.23*	.52
Text organization	← Text quality	.41*	.91	.37*	.85
	← Text quality	.21*	.46	.13*	.29

Note: \*indirect effects  
VD: Vocabulary depth

### *Development of vocabulary depth and text quality*

The results confirm the first two research hypotheses, as there was proven development in both vocabulary depth and text quality.

With regard to the developmental path of the vocabulary depth variable, we found that changes took place in the responses which implied better knowledge of the operations involved in the elicitation of synonyms and antonyms at higher grades. These results confirm the developmental increase in lexicon asserted by Berman and Nir-Sagiv (2009) and reveal that, as Nagy and Scott (2000) stated, vocabulary knowledge develops incrementally: it usually starts incomplete but gradually approximates adult knowledge (Clark, 1995). It cannot be explained as all-or-nothing but instead involves gradual acquisition. Therefore, these results do not match those of Kieffer and Lesaux (2012), who found no differences, although this could be because they conducted an in-depth evaluation of lexis with a multiple-choice task. The choice of different tasks reveals different ways of conceiving lexical depth (Schmitt, 2014): as a capacity more closely related to lexical fluency, that is, with the skill of accessing the vocabulary in one's own lexicon, as in their study, or as a capacity related to knowledge of a set of different aspects, as in our study.

In terms of text quality, we found that there are differences in the variables observed between the texts produced by children in the first grade and those produced by children in the fourth grade, as hypothesized. As the grades get higher, the texts produced are significantly better organized, which means that at the end of the early years of compulsory school many of the participants in the sample were capable of constructing descriptive texts with more prototypical components (introduction and conclusion and as series of descriptive clauses with support). Likewise, we also found a developmental increase in the number of words. That is, the texts written by the students in the fourth grade were longer. However, even though there were significant differences in subordination among grades, we could not clearly establish developmental differences in this variable. In short, we found that there are differences between the grades in all the variables and developmental gains in all except the subordination rate. Even though this last result seems inconsistent with the countless studies following in the footsteps of Hunt (1965), which have demonstrated that syntactic complexity in the production of written texts increases with age, it is in line with a series of studies which have proven that the effect of educational level interacts with the type of text produced. Wagner et al. (2011) found that the number of words, syntactic complexity and text organization explained the differences found between students from the first to fourth grades who wrote argumentative texts, but that syntactic complexity was the variable that explained these differences the least. Likewise, for descriptive texts, Beers and Nagy (2011) found no effect of the students' grade at school. These authors attributed this lack of effect to the fact that, just as in this study, the differences were studied in two-year intervals; however, in our opinion, it could be due to the fact that when description is integrated into other communicative functions, subordination is a syntactic resource that economizes and is therefore necessary: for example, in a story, 'Once upon a time a good teacher taught at a school in a small village that was in the bottom of a valley'. However, to say that someone is a good teacher, it is enough to use simple sentences: 'she's empathetic', 'he's friendly', 'she explains things well'.



### **Effect of vocabulary depth on text quality**

However, the most important finding of this study is confirmation of the effect of vocabulary depth on text quality in the early grades of primary school. Therefore, we confirmed the third research hypothesis: students with higher scores on vocabulary depth write longer texts with greater syntactic complexity and earn higher scores on text organization. According to Castillo and Tolchinsky (2018), who found equivalent results for Catalan, we believe that this is because establishing relationships of meaning among words helps learners free up cognitive effort so they can more effectively attend to the characteristics of the descriptive texts themselves. These results expand upon those found in previous studies (Olinghouse & Leaird, 2009; Olinghouse & Wilson, 2013), as they show that lexical aspects have to be considered not only as indicators but also as factors of influence in text quality.

### **Conclusions**

Even though this study did not analyse teaching-learning processes, a series of didactic implications can be drawn from the results.

On the one hand, we have a body of evidence that enables us to pinpoint some of the learning results of written expression which are formulated in a somewhat imprecise way, at least in the Spanish curriculum. For example, given that we found that descriptive texts are significantly better organized, where our regulations state that students should be able to 'write different kinds of texts, adapting the language to the textual features' as an evaluable learning standard in fourth grade (Section 3: Written communication. Writing), it could instead say 'write descriptive texts with the following prototypical components: introduction, description, support and conclusion'.

On the other hand, we found that there is a relationship between the developmental gains in vocabulary depth and the developmental gains in text quality which does enable us to venture that associating vocabulary instruction with the development of not only reading comprehension but also text composition could be an effective pedagogical practice.

Given the results of this study, one research line that could prove worthwhile in the future is an intervention study exploring the effects of an instructional process aimed at increasing vocabulary depth on the quality of both descriptive text and other types of texts.

## Profundidad de vocabulario y su contribución a la calidad textual en los primeros años de Educación Primaria

Uno de los objetivos principales de la Educación Primaria es que los escolares sean capaces de escribir, de forma efectiva, textos con distintas características en función de cuál sea su propósito comunicativo y, consecuentemente, se establecen una serie de resultados de aprendizaje que, a tal efecto, deben alcanzarse en cada nivel escolar. No obstante, el análisis del currículo español, por ejemplo, evidencia que no se establece cuál es la progresión prevista para la adquisición de los resultados de aprendizaje ni tampoco se aporta información sobre cuáles son los factores que, en esta etapa educativa, contribuyen a la calidad textual.

Se ha demostrado que el conocimiento de vocabulario que tienen los alumnos en primer curso de Educación Primaria explica más del 30% de la varianza de su nivel de comprensión de lectura en el último curso (Cunningham & Stanovich, 1997), pero ¿tiene este tipo de conocimiento el mismo poder explicativo respecto a la expresión escrita? Los distintos modelos que se han propuesto para explicar el proceso de construcción de un texto coinciden en que intervienen dos tipos de conocimiento: el del contenido y el lingüístico (léxico, sintáctico y retórico) (Tolchinsky, 2015). Para componer un texto sobre un tema determinado el escritor ha de saber seleccionar la información pertinente, es decir, ha de dominar el contenido, pero también debe saber seleccionar los términos más adecuados para la expresión de ese contenido y ha de conocer, por ejemplo, qué otros significados pueden tener dependiendo del contexto y cómo esos términos se asocian entre sí y con otros. Sin embargo, aunque es indudable que el conocimiento léxico es uno de los recursos fundamentales en la elaboración de un texto, la mayoría de los estudios se han centrado preferentemente en su relación con la comprensión lectora (Laufer & Ravenhorst-Kalovski, 2010; Ouellette & Beers, 2010; Sénéchal et al., 2006) y su potencial contribución a la calidad de los textos escritos está aún por explorar. Se ha visto, por otra parte, que, mientras la capacidad de descodificar está relacionada con el número de vocablos que una persona conoce (*amplitud de vocabulario*), la capacidad de acceder al significado de los textos depende de la profundidad con que haya adquirido su vocabulario, es decir, de su habilidad para manipular el significado de los vocablos conocidos (Ouellette, 2006) ¿Qué efecto cabe esperar de estas dos dimensiones del conocimiento léxico respecto a la producción de textos escritos?

El propósito de este estudio es analizar cómo evoluciona la *profundidad del vocabulario* en el alumnado de Educación Primaria y comprobar si tiene un efecto en la calidad de los textos que producen. Ello permitirá avanzar en la determinación de los rasgos de la calidad textual en esta población y, por lo tanto, en el establecimiento de unos resultados de aprendizaje realistas, lo que, a su vez, permitirá mejorar los procesos didácticos.

## Calidad textual: indicadores y habilidades involucradas

La descripción de los textos producidos por escolares es un cometido muy complicado y aún más complejo resulta analizar cómo evoluciona su calidad. Para el caso del español, se ha observado que los rasgos lingüísticos (como la complejidad sintáctica, la cohesión o el uso del vocabulario) explican las percepciones holísticas de calidad mejor que los rasgos de índole textual (Salas et al., 2016). No obstante, es necesario ir más allá de la evaluación holística, realizada por jueces externos, y abordar una evaluación analítica de rasgos característicos de los distintos niveles lingüísticos (palabra, oración, texto) (Verheyden et al., 2010).

Entre los numerosos indicadores que han sido utilizados en diferentes investigaciones sobre la calidad textual de textos producidos por estudiantes de Educación Primaria se encuentran rasgos lingüísticos como la riqueza léxica (Malvern et al., 2004) y rasgos de organización textual (Breetvelt et al., 1994; Tolchinsky, 2019).

Estas características textuales resultan de la combinación de distintas habilidades del sujeto-escritor, ya sean de expresión lingüística local, como, por ejemplo, las que tienen que ver con la selección léxica o la sintaxis oracional o habilidades discursivas más globales como, por ejemplo, la capacidad de estructurar bien un texto (Berman & Nir-Sagiv, 2009). Unas y otras se han utilizado para intentar explicar la calidad textual, pero los resultados obtenidos no son coincidentes (para un meta-análisis, ver Kent & Wanzek, 2016) y solo dan cuenta de un tercio de las diferencias encontradas (Salas & Tolchinsky, 2017). Así, por ejemplo, Abbott et al. (2010) señalan que el nivel de dominio de la ortografía (*spelling*) es la variable que mejor explica las diferencias longitudinales respecto a la composición de textos en los primeros siete cursos de la Educación Primaria en inglés. Sin embargo, Wagner et al. (2011) encontraron que, en la explicación de las diferencias evolutivas entre alumnos de 1er y 4º curso de Educación Primaria, también hablantes de inglés, la habilidad del sujeto con más poder explicativo era la fluidez de escritura a mano mientras que el dominio de la ortografía era menos explicativo. Por lo que respecta al español, en un estudio realizado sobre la misma muestra que el presente, Tolchinsky (2019) encontró que la capacidad de producir un discurso oral autosostenido, de manejar las relaciones de significado entre las palabras y una buena memoria de trabajo explicaban las diferencias en los textos descriptivos entre los grupos, mientras que los cambios de desarrollo en la estructura textual eran explicados por un alto dominio de la ortografía.

## Conocimiento del vocabulario

Durante el proceso de desarrollo léxico, las estrategias que los niños utilizan para aprender los significados de las palabras y relacionarlos entre sí cambian a medida que su representación interna del lenguaje crece y se reorganiza (Alexander Pan & Uccelli, 2010). Pasados los primeros años de vida, los niños siguen aprendiendo palabras nuevas, pero comienzan a aprender otros significados de las palabras que conocen y empiezan a establecer conexiones entre ellas. Es decir, al tiempo que su vocabulario se amplía, gana en profundidad. Consecuentemente, más allá de estudiar cuántas palabras conocen los aprendices y cómo se va incrementando su vocabulario a medida que progresan en su trayectoria escolar, es necesario averiguar qué relaciones

establecen entre palabras y cómo esas relaciones cambian con el tiempo, puesto que también es una cuestión de importancia para su progreso académico (Verhallen & Schoonen, 1998).

Para evaluar la amplitud de vocabulario (*vocabulary breadth*), o el número de entradas léxicas que un aprendiz es capaz de utilizar (Castillo & Tolchinsky, 2018), habitualmente se recurre a pruebas estandarizadas de vocabulario expresivo y/o receptivo. En cambio, evaluar la profundidad de vocabulario (*vocabulary depth*), o el grado de conocimiento que un aprendiz tiene de los múltiples significados de las palabras y de sus asociaciones (Alexander Pan & Ucelli, 2010), resulta mucho más complicado ya que se trata de una dimensión del desarrollo léxico que puede entenderse como el dominio de uno o varios de los siguientes tipos de conocimientos y habilidades que se solapan e interrelacionan (Ordóñez et al., 2002; Schmitt, 2014): conocimiento de los significados de palabras polisémicas, de sinónimos y antónimos, de palabras de la misma familia de palabras, de las colocaciones semánticas (las palabras o tipos de palabras que pueden y/o deben aparecer con un determinado vocablo), fluidez léxica (la capacidad de acceder al propio lexicón de manera rápida y precisa); y organización léxica (la habilidad de reconocer y proporcionar asociaciones de palabras).

En consecuencia, los estudios existentes sobre la adquisición de la profundidad léxica son escasos y poco coincidentes. Por ejemplo, los resultados del estudio de Kieffer y Lesaux (2012) con hablantes de inglés, como primera lengua o segunda lengua, de sexto grado de Educación Primaria, sugieren que la distinción entre amplitud y profundidad de vocabulario no refleja con precisión las diferencias individuales en ninguno de los aprendices. No obstante, Schwartz y Katzir (2012) detectaron que, tras un año de escolarización, los resultados de un grupo de estudiantes bilingües ruso-hebreos de segunda generación, de siete y ocho años, con un contexto socioeducativo favorable, mejoraron tanto la amplitud como la profundidad de vocabulario. Así pues, resulta necesario continuar investigando sobre este aspecto no solo para determinar su peso respecto a la amplitud de vocabulario sino también para averiguar cómo varía a través del tiempo, es decir, desde una perspectiva longitudinal.

## Relación entre vocabulario y calidad textual

Cuando planifica, textualiza o revisa un texto, el escritor tiene que ir eligiendo, entre los vocablos que conoce, los más apropiados en función del significado. Por eso, existen una serie de aspectos relacionados con la amplitud de vocabulario que han sido manejados en diversos estudios como indicadores de la calidad de los textos. Entre ellos, la diversidad léxica, es decir, el número de palabras diferentes respecto al número total de palabras, ha sido el rasgo más analizado y se ha comprobado que es la medida léxica que mejor explica las diferencias evolutivas entre alumnos de 2º y 4º grado de inglés (Olinghouse & Leaird, 2009). No obstante, también se sabe, que, cuando se trata de explicar la variación de la calidad en función del tipo de texto, entran en juego otros aspectos de la amplitud del vocabulario: para los textos persuasivos, el número de palabras temáticas, es decir, el número de palabras relacionadas con el tema sobre el que se escribe; para los textos informativos, el número de palabras temáticas y la madurez léxica, esto es, el número de palabras utilizadas que no se encuentran entre las más frecuentemente utilizadas en cada idioma (Olinghouse & Wilson, 2013).

Por otra parte, durante el proceso de escritura de un texto, el escritor ha de ser capaz de establecer una serie de relaciones semánticas, morfológicas y sintácticas entre las palabras que integran su lexicón (Castillo & Tolchinsky, 2018). Sin embargo, estos aspectos, que tienen que ver con la profundidad de vocabulario, no suelen ser incluidos entre los indicadores de la calidad de la expresión escrita. Son muy escasos los estudios que, como el de Gómez-Vera et al. (2016), realizado con alumnos chilenos de 4º grado, han demostrado que los rasgos de profundidad de vocabulario explican la variación en la calidad mejor que la diversidad léxica y que la sofisticación.

La contribución del conocimiento del vocabulario a la calidad de los textos escritos tampoco ha sido muy explorada, si bien los resultados obtenidos sugieren, de nuevo, que su efecto es diferente según este conocimiento se maneje en términos de amplitud o en términos de profundidad. En dos estudios realizados con niños hablantes de turco de 1º a 5º curso de Educación Primaria, Babayiğit y Stainthorp (2010, 2011) concluyeron que, junto con la memoria de trabajo, el vocabulario, medido en términos de amplitud, era el único predictor longitudinal fiable de la calidad del contenido de los textos, aunque no apreciaron ninguna contribución de esta variable a la calidad estructural de los mismos. No obstante, dos estudios realizados sobre textos producidos por alumnos de Educación Primaria en catalán (Castillo & Tolchinsky, 2018) y español (Tolchinsky, 2019), respectivamente, concluyeron que la profundidad de vocabulario, medida como conocimiento de sinónimos y antónimos, mejoraba con el nivel educativo y contribuía positivamente a la estructuración textual. Estos últimos resultados parecen indicar que la profundidad de vocabulario es una variable que tiene más potencialidad explicativa manejada como factor predictor que como indicador. El conocimiento de sinónimos y antónimos es una manifestación de las relaciones que el sujeto-escritor establece a nivel paradigmático, es decir, entre palabras que pueden o no aparecer en el mismo lugar, a diferencia de lo que sucede con otras dimensiones como el vocabulario temático o la diversidad léxica, que dan cuenta de las relaciones de significado que el sujeto-escritor establece entre las palabras que aparecen dentro del propio texto.

### El presente estudio

Aun no siendo exhaustivo, un repaso de los estudios que han abordado la calidad de los textos producidos por el alumnado de Educación Primaria permite componer un marco teórico en el que faltan estudios que concreten cómo evoluciona, en español, a lo largo de esta etapa evolutiva y que determinen con qué habilidades de los sujetos está más vinculada. De la revisión de la literatura se concluye también que, dentro de dichas habilidades, la profundidad de vocabulario, medida como conocimiento de sinónimos y antónimos, tiene un fuerte poder explicativo como factor predictor, si bien son necesarias más evidencias. Así pues, los objetivos que se plantearon para este estudio fueron dos:

- (1) Explorar cómo evoluciona el nivel de profundidad de vocabulario y la calidad de los textos producidos durante los primeros cursos de Educación Primaria, esta última a través de indicadores de productividad (o número de palabras), complejidad sintáctica y estructura textual.
- (2) Comprobar si el nivel de profundidad de vocabulario tiene efecto sobre la calidad de los textos producidos.

Las hipótesis que guiaron la investigación fueron:

- (1) El nivel de profundidad de vocabulario, concretado en el uso de sinónimos y antónimos, mejora con el nivel educativo.
- (2) A medida que se avanza en el nivel educativo, los textos estarán mejor estructurados, tendrán más palabras y serán sintácticamente más complejos.
- (3) Aquellos alumnos que tengan una mayor profundidad de vocabulario, concretada en el uso de sinónimos y antónimos, escribirán textos más largos, mejor estructurados y sintácticamente más complejos.

## Método

### Diseño

Para la consecución de estos objetivos, se optó por una metodología descriptiva del desarrollo combinando un estudio longitudinal con un estudio transversal. Puesto que no se trataba de un estudio de intervención, no se contempló como variable el proceso de enseñanza-aprendizaje.

### Participantes

En el estudio participó una muestra intencional de 332 estudiantes hablantes monolingües de español de siete escuelas de cuatro ciudades de España (Almería, Cádiz, Ciudad Real y Valencia) y se recogió información en dos aplicaciones separadas por un año. El primer año participaron 164 alumnos (45.1% niñas) de 1er curso de Educación Primaria ( $M = 6.68$ ) y 168 alumnos (48.8% niñas) de 3er curso de Educación Primaria ( $M = 8.71$ ). En la segunda aplicación los estudiantes mencionados se encontraban en 2º y 4º curso de Educación Primaria respectivamente.

### Tareas e instrumentos

Las tareas realizadas para recoger información sobre profundidad de vocabulario y calidad textual se realizaron en las dos aplicaciones, de forma idéntica.

### Profundidad de vocabulario

Puesto que, para la evaluación de la calidad textual se les iba a pedir a los participantes que produjeran varios textos descriptivos y supuestamente ello implicaba utilizar adecuadamente un buen número de adjetivos, se decidió que la evaluación de la profundidad de vocabulario se llevara a cabo mediante una tarea consistente en proporcionar un sinónimo o un antónimo para sendos conjuntos de adjetivos. Los adjetivos fueron seleccionados, con un criterio de frecuencia, del corpus CesCa y se tradujeron al español pues se trata de un corpus de textos escolares en catalán (Llaurado et al., 2012). Perteneían al mismo campo semántico (rasgos de personalidad).

Además, se optó por una tarea de producción oral y no por una de elección múltiple pues la profundidad de vocabulario debe concebirse como un continuum que abarca desde una cierta familiaridad con una palabra hasta la propiedad total de los diversos significados y la capacidad asociada para usarlos adecuadamente en contextos diferentes (Kieffer & Lesaux, 2012). Por otra parte, se entendió que esta tarea implicaba para los participantes la misma habilidad léxica involucrada en la producción textual, esto es, seleccionar palabras, no de un conjunto de alternativas ofrecidas, sino de su léxico mental.

Las investigadoras presentaron a cada uno de los participantes un mismo inventario de 28 palabras, una por una, y estos tuvieron que producir oralmente un sinónimo. A continuación, con el mismo procedimiento, presentaron otro inventario de 28 palabras, el mismo para cada uno de los participantes, y estos tuvieron que producir oralmente un antónimo. Las consignas utilizadas fueron las siguientes: ‘Ahora te voy a decir una serie de palabras y tú tendrás que decirme un sinónimo para cada una de ellas, es decir, una palabra que signifique lo mismo; ahora te voy a decir una serie de palabras y tú tendrás que decirme un antónimo para cada una de ellas, es decir, una palabra que signifique lo contrario’. Para asegurar la comprensión, se ejemplificó con tres sinónimos y con tres antónimos. La tarea se llevó a cabo individualmente en una sala diferente al aula donde recibían las clases y no tuvieron restricciones de tiempo, evitándose en todo momento que tuvieran que leer o escribir.

### **Calidad textual**

Las tareas de producción textual se realizaron después de las de evaluación de la profundidad de vocabulario. En la primera aplicación, cada participante escribió tres textos descriptivos: ‘cómo es un buen maestro’, ‘cómo es la casa de tus sueños’ y ‘cómo es una situación peligrosa’. Por su parte, los textos que se solicitó que escribieran en la segunda aplicación fueron: ‘cómo es tu mejor amigo/a’ y ‘cómo es tu lugar ideal de vacaciones’. Se optó por la producción de textos descriptivos por dos razones: porque es el único tipo de texto que aparece en todos los currículos de Educación Primaria de las distintas regiones de España como estándar de aprendizaje evaluable y porque, aun teniendo características estructurales propias, también puede aparecer integrado dentro de los otros tipos de texto (por ejemplo, el narrativo) que el alumnado de esta etapa debe aprender. Los participantes trabajaron al mismo tiempo, en su aula habitual, pero los textos fueron escritos individualmente, no como producto de una tarea de aprendizaje sino como producto de una tarea de evaluación, cada uno de ellos en tiempo máximo de 20 minutos y con no más de dos semanas de lapso entre unos y otros. En total, los participantes compusieron 492 textos en la primera aplicación y 336 en la segunda que fueron digitalizados.

### **Procedimiento de codificación y análisis**

Previo al análisis, se llevó a cabo un trabajo de codificación de las respuestas obtenidas en la tarea de profundidad de vocabulario y de los textos producidos por los participantes.

Para la primera codificación el equipo de investigación creó un sistema de 8 categorías basado en la combinación de cuatro dimensiones, que reproducen algunas de las operaciones implicadas en la elicitación de sinónimos y antónimos y que también fue aplicado en Castillo y Tolchinsky (2018) y Tolchinsky (2019):

### *Cambio de lexema*

el participante proporciona una palabra diferente a la palabra dada.

### *Ajuste al sentido de significado solicitado*

el participante proporciona una palabra con significado equivalente cuando se le solicita un sinónimo o una palabra con significado contrario cuando se le solicita un antónimo.

### *Preservación de la categoría gramatical*

el participante proporciona un adjetivo.

### *Convencionalidad*

el participante proporciona una palabra que aparece como sinónimo o antónimo en cuatro diccionarios tomados como referencia.

A su vez, y a fin de poder obtener una calificación del desempeño en esta tarea y, por lo tanto, del nivel de profundidad de vocabulario, se estableció un sistema de puntuación del 1 al 5, en función de cuántas y qué dimensiones se manifestaran en cada respuesta: cambio de lexema, 1 punto; ajuste al sentido de significado solicitado, 2 puntos, preservación de la categoría gramatical, 1 punto y convencionalidad, 1 punto.

Así, si, cuando se les solicitaba un sinónimo o un antónimo, los participantes ofrecían como respuesta un adjetivo diferente, con significado equivalente o contrario, respectivamente, y que aparecía consignada como sinónimo o antónimo en los diccionarios consultados, dicha respuesta se incluía dentro de la categoría 8 (ver material complementario): por ejemplo, *contento* como sinónimo de *alegre*, o *callado* como antónimo de *charlatán*. Las respuestas de esta categoría se puntuaron con un 5 porque reflejaban un máximo conocimiento de las operaciones semánticas requeridas para establecer relaciones de antonimia y sinonimia (ajuste al sentido de significado solicitado), conocimiento léxico (cambio de lexema y convencionalidad) y conocimiento gramatical (preservación de categoría gramatical).

Si, en cambio, cuando se les solicitaba un sinónimo o un antónimo, los participantes ofrecían adjetivos diferentes, con significado equivalente o contrario, respectivamente, y pero que no aparecían consignadas como sinónimos o antónimos en los diccionarios consultados, por ejemplo, *feliz* como sinónimo de *alegre* o *silencioso* como antónimo de *callado*, las respuestas se incluían en la categoría 7. Las respuestas de esta categoría se puntuaron con un 4 porque de ellas se podía deducir un buen conocimiento semántico y gramatical si bien no evidenciaban conocimiento de las asociaciones léxicas convencionales recogidas en los diccionarios. Asimismo, dada la frecuencia con la que aparecían, se consideró que debían incluirse en esta categoría aquellas respuestas que, cumpliendo con las características antes citadas: (a) eran resultado de una metáfora o comparación — por ejemplo, *fideo* como sinónimo de *delgado* o *enano* como antónimo de *alto*; (b) podían considerarse válidas en un determinado registro



o variedad — por ejemplo, *machote* como sinónimo de *fuerte* o *cagao* como antónimo de *valiente* o; (c) podían considerarse válidas en un contexto infantil — por ejemplo, *bueno* como sinónimo de *listo* (como consecuencia de la frecuente valoración de los adultos ‘es un niño muy bueno y muy listo’) o *enfadado* como antónimo de *guapo* (como consecuencia de la frecuente apreciación de los adultos ‘cuando te enfadas te pones muy feo’).

En el resto de categorías se incluyeron las respuestas en las que dos o más dimensiones no estaban presentes.

Cuando a la solicitud de un sinónimo o de un antónimo respondían con una palabra diferente que reflejaba el significado en el sentido solicitado pero que no era un adjetivo y, consecuentemente, no podía aparecer en los diccionarios como tal, las respuestas se incluyeron en la categoría 6. Así, por ejemplo, cuando dijeron *miedo* como sinónimo de *cobarde* o *pesimismo* como antónimo de *optimista*. Las respuestas de esta categoría se puntuaron con un 4 porque, como las respuestas de la categoría 7, reflejaban conocimiento léxico y semántico, aunque menos conocimiento gramatical.

Cuando, como respuesta, ofrecían una palabra diferente que reflejaba el significado en el sentido solicitado pero que no podía considerarse que perteneciese al español por incluir determinados cambios morfológicos, por ejemplo, ‘*hablatón*’ como sinónimo de *charlatán* o ‘*calentero*’ como antónimo de *friolero*, las aportaciones de los participantes se incluyeron en la categoría 5 y se puntuaron con un 3.

Si, al solicitarles un sinónimo y antónimo los participantes aportaban respuestas coherentes con el sentido de significado solicitado pero que no eran adjetivos sino sintagmas o frases, dichas respuestas se incluyeron en la categoría 4 y se puntuaron con un 3: por ejemplo, *no lo sabe hacer* como sinónimo de *inseguro* o *echarse para atrás* como antónimo de *decidido*. En cambio, cuando aportaban sintagmas o frases, pero incluían la palabra por la que se estaba preguntando, por ejemplo, *llora mucho* como sinónimo de *llorón* o *que no es aburrido* como antónimo de *aburrido*, las respuestas se incluyeron en la categoría 3 y se puntuaron con un 2 pues la única de las cuatro dimensiones que manifestaban era el ajuste al sentido de significado solicitado.

Aquellas respuestas de los participantes que, teniendo un significado equivalente o contrario, según se requiriese, no suponían un cambio de lexema, pero sí algún tipo de variación morfológica, normativa o no, se incluyeron en la categoría 2 y se puntuaron con un 3 pues se consideró que reflejaban un mayor conocimiento gramatical determinado por la preservación de la categoría gramatical: por ejemplo, *rabioso* como sinónimo de *rabioso* o ‘*intriste*’ como antónimo de *triste*. Finalmente, en la categoría 1 se incluyeron todas las respuestas consideradas erróneas por no ajustarse al sentido de significado solicitado: por ejemplo, *enfadada* como sinónimo de *alegre* o *malón* como antónimo de *malo*.

Así, el puntaje máximo que se podía obtener fue de 280 puntos (28 sinónimos x 5 puntos de puntuación máxima para cada uno; 28 antónimos x 5 puntos de puntuación máxima para cada uno). Antes de proceder a la codificación de toda la muestra, se llevó a cabo un proceso de validación interna, que se realizó mediante la valoración del acuerdo interjueces. El grado de acuerdo entre los tres jueces expertos fue superior al 90% en todos los casos.

Por otra parte, los textos producidos por los participantes fueron codificados en función de los siguientes aspectos:

### Productividad

Número de palabras gráficas, es decir, el número de cadenas de letras delimitadas por un espacio en blanco a cada lado.

### Índice de subordinación

Número de cláusulas subordinadas exceptuando las completivas respecto al total de cláusulas.

### Estructura textual

Presencia o ausencia (1/0) de los componentes funcionales de los textos descriptivos: cláusulas de introducción, descriptivas, justificativas y de conclusión. La calificación se realizó en cada texto y posteriormente se calculó la media aritmética para cada aplicación. Se utilizó una escala de 0 a 8 según se describe a continuación (Tabla 1):

Para responder a la primera de las preguntas de investigación (objetivo 1), es decir, para valorar los cambios evolutivos en las diferentes variables, se analizó el efecto del curso. Se realizaron pruebas de análisis de varianza simple (ANOVA) cuando el análisis era entre diferentes sujetos, y pruebas *t* (de muestras relacionadas) cuando las diferencias objeto de análisis se referían a las diferentes aplicaciones con los mismos sujetos.

Para responder a la segunda pregunta de investigación (objetivo 2), es decir, para explicar las relaciones entre las variables seleccionadas, se comprobó un modelo que se describe en la Figura 1. Se hipotetizó la existencia de una variable latente denominada *calidad textual* sobre la que tendría efecto la variable observada *profundidad de vocabulario* y que estaría explicada por tres variables observadas: *número de palabras*, *índice de subordinación* y *estructura textual*.

Se comprobó la estructura propuesta mediante un modelo de ecuaciones estructurales y se analizó si esta estructura explicativa es equivalente a lo largo de los cursos mediante un análisis multigrupo. Para evitar los efectos del aprendizaje de prueba se utilizan sólo las respuestas de la primera aplicación (1er y 3er curso).

tabla 1. Sistema de evaluación y calificación de la estructura textual.

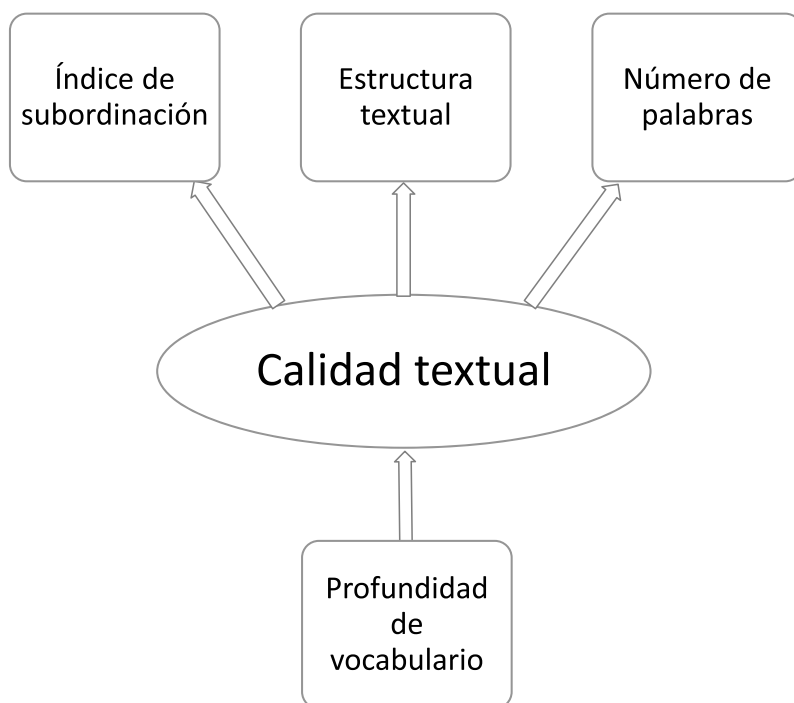
Calificación	Introducción	Descriptivas	Justificativas	Conclusión
8	1	1	1	1
7	0	1	1	1
6	1	1	1	0
5	0	1	1	0
4	0	1	0	1
3	1	1	0	0
2	0	1	0	0
1*	1	0	0	0
	0	0	1	0
	0	0	0	1
0	0	0	0	0

Nota:

0 No aparece

1 Aparece

1\*ausencia de cláusulas descriptivas y presencia de alguna de los otros tres tipos



**Figura 1.** Modelo propuesto para la explicación de la calidad textual y su relación con la profundidad de vocabulario

La comparación entre grupos se estableció progresivamente, siguiendo las indicaciones de Byrne (2004), y se analizaron los índices de ajuste para cada uno de los modelos propuestos. En el primer modelo se estima la invarianza factorial sin restricciones, asumiendo diferencias en los grupos que componen la muestra. El segundo modelo restringe las saturaciones (*measurement weights*) igualándolas en todos los grupos, lo que supone considerar que los pesos de cada ítem por factor son equivalentes. El tercer modelo añade la restricción en el efecto de los pesos estructurales (*structural weights*), dicho de otro modo, se iguala el efecto de la profundidad de vocabulario sobre la calidad textual. El cuarto modelo añade la restricción en las covarianzas estructurales (*structural covariances*), asumiendo igualdad en las varianzas y covarianzas dentro de los grupos. Por último, el quinto modelo añade la restricción en los residuos de medida (*measurement residuals*), asumiendo la equivalencia de los errores.

Para determinar la bondad de ajuste se utilizaron los siguientes índices y criterios: Chi cuadrado ( $\chi^2$ ) mayor que .05; razón de Chi cuadrado entre grados de libertad ( $\chi^2/df$ ), menor que 3; error cuadrático medio de aproximación (*RMSEA*) menor que .05; índice de bondad de ajuste comparativo (*CFI*), índice Tucker-Lewis (*TLI*) e índice de ajuste normalizado (*NFI*) valores superiores a .95.

## Resultados

### *Evolución de la profundidad de vocabulario*

Tal como se observa en la [Tabla 2](#), las respuestas a la tarea de profundidad semántica variaron cuando se incrementó el curso. Así, se fueron reduciendo progresivamente las respuestas de las categorías 1, 2, 3 y 4, que suponen no saber aplicar o aplicar parcialmente las distintas operaciones implicadas en la elicitación de un sinónimo o un antónimo: seleccionar una palabra diferente, que también fuese adjetivo y con un significado equivalente o contrario, respectivamente,

Por otro lado, se incrementaron progresivamente las respuestas correctas, es decir, las categorías 7 y 8, que implican el dominio de estas operaciones así como, en el caso de las respuestas de la categoría 8, un conocimiento de las relaciones de significado convencionales, esto es, las recogidas en los diccionarios.

Las categorías 5 y 6 mantienen un valor residual a lo largo de los cuatro cursos.

Estas diferencias evolutivas en profundidad de vocabulario (entre los cuatro grupos) son significativas tanto para los sinónimos (Chi-cuadrado = 1,027.87;  $df = 21$ ; sig. < .01) como para los antónimos (Chi-cuadrado = 1,229.83;  $df = 21$ ; sig. < .01).

### *Evolución de los indicadores de calidad textual*

La [Tabla 3](#) describe las diferencias en las puntuaciones medias de las calificaciones de estructura textual, número de palabras e índice de subordinación en cada una de las aplicaciones de cada uno de los textos. Como se observa, el efecto del curso fue significativo (< .01) a favor del curso más alto en todas las variables y todos los textos, tanto en la primera como en la segunda aplicación. Los resultados obtenidos permiten comprobar que, en la muestra participante, todos los indicadores seleccionados para valorar la calidad de la producción textual fueron aumentando progresivamente cuando se incrementó el curso académico.

**Tabla 2.** Porcentaje de respuesta para cada una de las ocho categorías de sinónimos y antónimos en función del curso.

		Curso			
		Primera cohorte		Segunda cohorte	
		1º	2º	3º	4º
Categorías Sinónimos	1	10.6%	8.8%	4.7%	5.1%
	2	11.3%	10.1%	7.6%	4.2%
	3	41.7%	34.6%	35.3%	23.4%
	4	2.0%	1.0%	1.3%	0.3%
	5	0.3%	0.2%	0.7%	0.5%
	6	1.4%	1.8%	1.6%	1.5%
	7	10.9%	14.9%	15.3%	21.0%
	8	21.8%	28.6%	33.7%	43.9%
Categorías Antónimos	1	7.1%	6.1%	3.3%	2.7%
	2	3.0%	4.4%	3.8%	4.5%
	3	44.7%	28.3%	28.0%	14.4%
	4	0.6%	0.3%	0.5%	0.1%
	5	1.5%	1.8%	3.0%	2.6%
	6	0.8%	0.8%	0.8%	0.6%
	7	14.7%	20.6%	21.3%	26.7%
	8	27.4%	37.9%	39.3%	48.2%

**Tabla 3.** Efecto del curso en la estructura textual, el número de palabras, el índice de subordinación y la profundidad de vocabulario. ANOVA.

		<i>Estructura textual</i>					
		Primera aplicación			Segunda aplicación		
		Texto 1	Texto 2	Texto 3	Texto 1	Texto 2	Texto 3
1º		3.35	2.46	3.67	4.44	3.98	
3º		4.32	3.80	4.83	5.16	4.69	
Total		3.85	3.14	4.27	4.80	4.34	
		$F = 18.90; gl = 321; p < .01$			$F = 7.86; gl = 293; p < .01$		
		$F = 42.07; gl = 311; p < .01$			$F = 6.82; gl = 268; p < .01$		
		$F = 24.90; gl = 312; p < .01$			$F = 24.53; gl = 318; p < .01$		
		<i>Número de palabras</i>					
		Primera aplicación			Segunda aplicación		
		Texto 1	Texto 2	Texto 3	Texto 1	Texto 2	Texto 3
1º		26.48	32.09	26.60	60.27	41.50	
3º		67.32	72.76	59.24	89.71	61.56	
Total		47.34	52.93	42.63	74.99	51.62	
		$F = 162.19; gl = 322; p < .01$			$F = 35.69; gl = 313; p < .01$		
		$F = 111.06; gl = 321; p < .01$			$F = 24.53; gl = 318; p < .01$		
		$F = 147.99; gl = 325; p < .01$			$F = 24.53; gl = 318; p < .01$		
		<i>Índice de Subordinación</i>					
		Primera aplicación			Segunda aplicación		
		Texto 1	Texto 2	Texto 3	Texto 1	Texto 2	Texto 3
1º		29.91	19.61	30.42	30.50	33.95	
3º		39.12	29.67	38.18	39.12	35.12	
Total		34.66	24.65	34.36	34.99	34.54	
		$F = 10.81; gl = 319; p < .01$			$F = 9.41; gl = 316; p < .01$		
		$F = 13.30; gl = 306; p < .01$			$F = 0.14; gl = 316; p > .05$		
		$F = 6.93; gl = 324; p < .01$			$F = 0.14; gl = 316; p > .05$		

Para valorar los cambios evolutivos, se compararon las puntuaciones de cada participante en dos cursos académicos consecutivos. Los resultados de la prueba  $t$  para muestras relacionadas indicaron que se produjeron ganancias evolutivas estadísticamente significativas en estructura textual, tanto entre 1º y 2º [ $n = 156$ ,  $d.m. = -1.16$  (3.13–4.29), *i.c.* 95% (-1.51, -0.82),  $t = -6.70$ ,  $gl = 155$ ,  $sig < .01$ ] como entre 3º y 4º [ $n = 162$ ,  $d.m. = -0.57$  (4.35–4.92), *i.c.* 95% (-0.91, -0.24),  $t = -3.36$ ,  $gl = 161$ ,  $sig < .01$ ] y en el número de palabras, tanto entre 1º y 2º [ $n = 159$ ,  $d.m. = -22.76$  (27.74–50.50), *i.c.* 95% (-26.88, -18.63),  $t = -10.90$ ,  $gl = 158$ ,  $sig < .01$ ] como entre 3º y 4º [ $n = 162$ ,  $d.m. = -9.38$  (66.06–75.44), *i.c.* 95% (-14.54, -4.22),  $t = -3.59$ ,  $gl = 162$ ,  $sig < .01$ ].

Estos resultados indican que se produjeron cambios individuales en la calidad de las producciones escritas que se reflejaron en una mejor estructuración de los textos y un mayor número de palabras. Sin embargo, a pesar de que sí aparecieron diferencias entre las medias de los cursos en las aplicaciones para la variable subordinación y las puntuaciones se incrementaron con el curso, el análisis de las diferencias individuales en este indicador muestra que no son significativas ( $t = 0.80$ ,  $sig. = .42$ ).

### Efecto de la profundidad de vocabulario sobre la calidad textual

Se propone que las relaciones entre las variables descritas anteriormente responden a la estructura que se describe en la [Figura 1](#) y que hipotetiza:

- La existencia de la variable latente *calidad textual*.
- Esta variable latente estaría explicada por tres variables observadas: *número de palabras*, *índice de subordinación* y *estructura textual*.
- La variable observada profundidad de vocabulario tiene efecto directo sobre la variable latente calidad textual e indirecto sobre número de palabras, índice de subordinación y estructura textual.
- Existen diferencias en el modelo de relaciones entre calidad textual y profundidad de vocabulario en función del curso

Para comprobar estas hipótesis se realizó un análisis multigrupo. Los resultados de los datos de ajuste ([Tabla 4](#)) indicaron que el modelo sin restricciones obtuvo un buen ajuste para todos los indicadores analizados. La significatividad asociada a Chi cuadrado fue mayor de .05, el cociente entre este indicador y los grados de libertad fue menor que 3, *RMSEA* menor que .05 y *CFI*, *TLI* y *NFI* superiores a .95, esto indica que los datos recogidos se ajustaron satisfactoriamente al modelo explicativo propuesto.

Los resultados apoyaron la existencia de la variable latente calidad textual y confirmaron que está explicada por la puntuación en estructura textual y en menor

**Tabla 4.** Comparación de ajuste de modelos. Análisis multigrupo.

Modelo	$\chi^2$ (sig.)	$df$	$\chi^2/df$	<i>RMSEA</i>	<i>CFI</i>	<i>TLI</i>	<i>NFI</i>
Sin restricciones	5.93 (.204)	4	1.48	.038	.988	.939	.967
Invarianza métrica	18.06 (.006)	6	3.01	.078	.924	.746	.899
Invarianza factorial	18.07 (.012)	7	2.58	.069	.930	.800	.899
Covarianza estructural	19.50 (.012)	8	2.43	.066	.927	.818	.891
Valores residuales	97.69 (.000)	11	8.88	.155	.453	.005	.452

**Tabla 5.** Coeficientes de regresión estandarizados y efectos estandarizados totales, directos e indirectos. 1er y 3er curso.

		Efectos estandarizados			
		1º		3º	
		PV	Calidad	PV	Calidad
Calidad textual	← Prof. vocabulario	.45		.43	
Índice subordinación	← Calidad textual	.24*	.53	.23*	.52
Estructura textual	← Calidad textual	.41*	.91	.37*	.85
Número palabras	← Calidad textual	.21*	.46	.13*	.29

Nota: \*efectos indirectos

medida (aunque también significativamente) por el número de palabras y el índice de subordinación. También se confirmó el efecto de la profundidad de vocabulario sobre la calidad textual, con un efecto estandarizado directo superior a .04 (.45 en 1er curso; .43 en 3º), que se considera un valor adecuado para considerar el efecto de una variable sobre otra. Dicho de otro modo, los resultados indicaron que una mayor profundidad de vocabulario determina una mayor calidad en los textos producidos.

También se confirmó que los dos cursos presentan diferencias en el modelo de relaciones entre calidad textual y profundidad de vocabulario. Todos los índices de ajuste disminuyeron al igualar las saturaciones, los efectos estructurales, las covarianzas estructurales o los residuos de medida. Ninguna de estas alternativas alcanzó valores adecuados para un correcto ajuste.

La **Tabla 5** describe los coeficientes de regresión y los efectos estandarizados en los dos cursos conjuntamente para facilitar su comparación. Los resultados indican que al aumentar el curso disminuyó ligeramente el efecto directo de la profundidad de vocabulario sobre la calidad textual, así como el efecto indirecto de esta variable sobre los tres indicadores de calidad, especialmente sobre el número de palabras.

Por el contrario, parece que al aumentar el curso se consolidó el modelo de relaciones entre calidad textual y profundidad de vocabulario basado en las tres variables propuestas.

## Discusión

El propósito de este estudio era doble. Por un lado, trazar la evolución de un tipo de conocimiento lingüístico, el conocimiento léxico, medido en términos de profundidad de vocabulario, y la evolución de la calidad textual, medida a nivel de palabra (número de palabras), a nivel oracional (complejidad sintáctica) y a nivel textual (estructura textual), durante los primeros cursos de la Educación Primaria. Para ello, se analizaron los resultados obtenidos por los grupos de 1º, 2º, 3º y 4º y las diferencias en el paso de 1º a 2º y de 3º a 4º. Por otro lado, determinar si la profundidad de vocabulario es una variable que influye en el nivel de calidad de los textos producidos por el alumnado y si hay diferencias en este efecto en relación al curso. Para ello, se llevó a cabo un análisis de estructura de covarianzas multigrupo con los alumnos de 1er y 3er curso para controlar el efecto de la repetición de tareas.

### *Evolución de la profundidad del vocabulario y de la calidad textual*

Los resultados obtenidos confirman las dos primeras hipótesis de la investigación puesto que demuestran que hay una evolución contrastada tanto en profundidad de vocabulario como en calidad textual.

Respecto al recorrido evolutivo de la variable profundidad de vocabulario, se ha comprobado que se producen cambios en las respuestas que implican un mayor conocimiento de las operaciones implicadas en la elicitación de sinónimos y antónimos a medida que se incrementa el curso. Estos resultados confirman el incremento evolutivo que respecto al uso léxico señalaron Berman y Nir-Sagiv (2009) y evidencian que, como señalan Nagy y Scott (2000), el conocimiento del vocabulario se desarrolla por pasos: al principio normalmente es incompleto, pero se va aproximando gradualmente al conocimiento de los adultos (Clark, 1995). No se explica en términos de todo-o-nada, sino que se trata de una adquisición gradual. No coinciden, en cambio, con los de Kieffer y Lesaux (2012) que no encontraron diferencias, pero esto puede deberse a que evaluaron la profundidad léxica con una tarea de elección múltiple. La elección de distintas tareas revela distintas maneras de concebir la profundidad léxica (Schmitt, 2014): como una capacidad más relacionada con la fluidez léxica, es decir, con la habilidad para acceder al vocabulario del propio lexicón, en su caso, y, como una capacidad relacionada con el conocimiento de un conjunto de distintos aspectos, en nuestro estudio.

Respecto a la calidad textual, se comprobó que, como se hipotetizó, existen diferencias en las variables observadas entre los textos producidos por los niños de primer curso de Educación Primaria y los textos que producen los niños de 4º. A medida que se avanza de curso, los textos producidos están significativamente mejor estructurados lo que quiere decir que, al final de los primeros años de la escolarización obligatoria, muchos de los participantes en la muestra eran capaces de construir un texto descriptivo con más componentes prototípicos (introducción y conclusión y una serie de cláusulas descriptivas que, además, estaban justificadas). Igualmente, se observó un incremento evolutivo en el número de palabras. Es decir, los textos escritos por los alumnos de 4º fueron más largos. Sin embargo, aunque aparecieron diferencias significativas en la subordinación entre los cursos, no pudimos establecer claramente diferencias evolutivas en esta variable. En resumen, se comprueba que hay diferencias entre los cursos en todas las variables y ganancias evolutivas en todas excepto en el índice de subordinación. Si bien este último resultado resulta inconsistente con los numerosos estudios que, siguiendo la estela de Hunt (1965), han demostrado que, en la producción de textos escritos, la complejidad sintáctica aumenta con la edad, está en línea con otra serie de trabajos que han probado que el efecto del nivel educativo interacciona con el tipo de texto producido. Wagner et al. (2011) observaron que el número de palabras, la complejidad sintáctica y la organización textual explicaba las diferencias encontradas entre alumnos de 1º a 4º que escribieron un texto argumentativo, pero encontraron que la complejidad sintáctica era la variable que menos explicaba dichas diferencias. Por su parte, para el caso de los textos descriptivos Beers y Nagy (2011) tampoco encontraron un efecto del curso. Estos autores atribuyen esta falta de efecto a que, como en el presente estudio, las diferencias se estudian en intervalos de dos años pero, a nuestro juicio, podría deberse a que, cuando la descripción está integrada en otras funciones comunicativas, la subordinación es un recurso sintáctico que economiza y, por tanto, es un recurso necesario: por ejemplo, en una narración, ‘érase una vez un buen maestro que daba clase en una escuela de un pequeño



pueblo que estaba en el fondo de un valle'. Sin embargo, para decir que alguien es un buen maestro basta utilizar oraciones simples: 'es empático', 'es amable', 'explica bien'.

### **Efecto de la profundidad de vocabulario sobre la calidad textual**

No obstante, el mayor hallazgo de este estudio consiste en haber podido probar el efecto de la profundidad de vocabulario sobre la calidad de los textos en los primeros cursos de la Educación Primaria, comprobando, por tanto, la tercera de las hipótesis de investigación: aquellos estudiantes que obtienen mejores puntuaciones respecto a profundidad de vocabulario escriben textos más largos, con una mayor complejidad sintáctica y con una mayor puntuación en cuanto a la estructura textual. De acuerdo con Castillo y Tolchinsky (2018), que obtuvieron resultados equivalentes para el catalán, entendemos que esto es así porque establecer relaciones de significado entre palabras ayuda a los aprendices a liberar esfuerzo cognitivo y les permite atender más eficazmente a las características propias de los textos descriptivos. Estos resultados amplían los obtenidos en investigaciones anteriores (Olinghouse & Leaird, 2009; Olinghouse & Wilson, 2013) pues muestran que, en lo que respecta a la calidad textual, los aspectos léxicos han de ser contemplados no solo como indicadores sino también como factores de influencia.

### **Conclusiones**

Si bien en este estudio no se analizaron los procesos de enseñanza-aprendizaje, de los resultados obtenidos se pueden extraer una serie de implicaciones didácticas.

Por un lado, disponemos de un conjunto de evidencias que permiten concretar algunos de los resultados de aprendizaje de la expresión escrita que, al menos en el currículo español, están formulados de una forma un tanto imprecisa: así, por ejemplo, puesto que se ha visto que los textos descriptivos están significativamente mejor estructurados, donde nuestra normativa establece como estándar de aprendizaje evaluable de 4º curso (Bloque 3: Comunicación Escrita. Escribir) que el alumno 'escribe diferentes tipos de textos adecuando el lenguaje a las características textuales', se podría decir que 'escribe textos descriptivos con los siguientes componentes prototípicos: introducción, descripción, justificación y conclusión'.

Por otro lado, se ha probado que existe una relación entre las ganancias evolutivas respecto a profundidad de vocabulario y las ganancias evolutivas respecto a calidad textual que sí permite aventurar que vincular la instrucción del vocabulario no solo con el desarrollo de la comprensión lectora sino también con el aprendizaje de la composición textual puede resultar una práctica pedagógica efectiva.

Vistos los resultados de este estudio, una línea de investigación que sería interesante abordar en el futuro es un estudio de intervención en el que se exploren los efectos de un proceso de instrucción dirigido a incrementar la profundidad de vocabulario sobre la calidad tanto de los textos descriptivos como de otro tipo de textos.

## Acknowledgements / Agradecimientos

This work was supported by the Ministerio de Economía y Competitividad [(EDU2012-36577)]. / *Este trabajo ha recibido el apoyo del Ministerio de Economía y Competitividad [(EDU2012-36577)].*

## Disclosure statement / Conflicto de intereses

No potential conflict of interest was reported by the authors. / *Los autores no han referido ningún potencial conflicto de interés en relación con este artículo.*

## ORCID

M. Dolores Alonso-Cortés-Fradejas  <http://orcid.org/0000-0001-8214-8188>

Mercedes López-Aguado  <http://orcid.org/0000-0002-4822-6901>

M. Teresa Llamazares-Prieto  <http://orcid.org/0000-0002-2188-1412>

## References / Referencias

- Abbott, R. D., Berninger, V. W., & Fayol, M. (2010). Longitudinal relationships of levels of language in writing and between writing and reading in grades 1 to 7. *Journal of Educational Psychology, 102*(2), 281–298. <https://doi.org/10.1037/a0019318>
- Alexander Pan, B., & Uccelli, P. (2010). Desarrollo semántico: El aprendizaje del significado de las palabras. In J. Berko Gleason, & N. Bernstein Ratner (Eds.), *Desarrollo del lenguaje* (7 ed., pp. 109–141). Pearson Educación.
- Babayigit, S., & Stainthorp, R. (2010). Component processes of early reading, spelling, and narrative writing skills in Turkish: A longitudinal study. *Reading and Writing, 23*(5), 539–568. <https://doi.org/10.1007/s11145-009-9173-y>
- Babayigit, S., & Stainthorp, R. (2011). Modeling the relationships between cognitive-linguistic skills and literacy skills: New insights from a transparent orthography. *Journal of Educational Psychology, 103*(1), 169–189. <https://doi.org/10.1037/a0021671>
- Beers, S. F., & Nagy, W. E. (2011). Writing development in four genres from grades three to seven: Syntactic complexity and genre differentiation. *Reading and Writing, 24*(2), 183–202. <https://doi.org/10.1007/s11145-010-9264-9>
- Berman, R., & Nir-Sagiv, B. (2009). Cognitive and linguistic factors in evaluating text quality: Global versus local. In V. Evans, & S. Pourcel (Eds.), *New directions in cognitive linguistics* (pp. 421–440). John Benjamins.
- Breetvelt, I., Van den Bergh, H., & Rijlaarsdam, G. (1994). Relations between writing processes and text quality: When and how? *Cognition and Instruction, 12*(2), 103–123. [https://doi.org/10.1207/s1532690xci1202\\_2](https://doi.org/10.1207/s1532690xci1202_2)
- Byrne, B. M. (2004). Testing for multigroup invariance using AMOS graphics: A road less traveled. *Structural Equation Modeling: A Multidisciplinary Journal, 11*(2), 272–300. [https://doi.org/10.1207/s15328007sem1102\\_8](https://doi.org/10.1207/s15328007sem1102_8)
- Castillo, C., & Tolchinsky, L. (2018). The contribution of vocabulary knowledge and semantic orthographic fluency to text quality through elementary school in Catalan. *Reading and Writing, 31*(2), 293–323. <https://doi.org/10.1007/s11145-017-9786-5>
- Clark, E. V. (1995). *The lexicon in acquisition* (Vol. 65). Cambridge University Press.
- Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33*(6), 934–945. <https://doi.org/10.1037/0012-1649.33.6.934>
- Gómez-Vera, G., Sotomayor, C., Bedwell, P., Domínguez, A. M., & Jéldrez, E. (2016). Analysis of lexical quality and its relation to writing quality for 4th grade, primary school students in Chile. *Reading and Writing, 29*(7), 1317–1336. <https://doi.org/10.1007/s11145-016-9637-9>

- Hunt, K. W. (1965). *Grammatical structures written in three grade levels. Research report no. 3.* National Council of Teachers of English.
- Kent, S. C., & Wanzek, J. (2016). The relationship between component skills and writing quality and production across developmental levels: A meta-analysis of the last 25 years. *Review of Educational Research*, 86(2), 570–601. <https://doi.org/10.3102/0034654315619491>
- Kieffer, M. J., & Lesaux, N. K. (2012). Knowledge of words, knowledge about words: Dimensions of vocabulary in first and second language learners in sixth grade. *Reading and Writing*, 25(2), 347–373. <https://doi.org/10.1007/s11145-010-9272-9>
- Lauffer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15–30.
- Llaurado, A., Martí, M. A., & Tolchinsky, L. (2012). Corpus CesCa: Compiling a corpus of written Catalan produced by school children. *International Journal of Corpus Linguistics*, 17(3), 428–441. <https://doi.org/10.1075/ijcl.17.3.06lla>
- Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development.* Palgrave Macmillan.
- Nagy, W. E., & Scott, J. A. (2000). Vocabulary processes. In M. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr, (Eds.), *Handbook of reading research* (Vol. 3, pp. 269–284). Lawrence Erlbaum Associates.
- Olinghouse, N. G., & Leaird, J. T. (2009). The relationship between measures of vocabulary and narrative writing quality in second-and fourth-grade students. *Reading and Writing*, 22(5), 545–565. <https://doi.org/10.1007/s11145-008-9124-z>
- Olinghouse, N. G., & Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing*, 26(1), 45–65. <https://doi.org/10.1007/s11145-012-9392-5>
- Ordóñez, C. L., Carlo, M. S., Snow, C. E., & McLaughlin, B. (2002). Depth and breadth of vocabulary in two languages: Which vocabulary skills transfer? *Journal of Educational Psychology*, 94(4), 719–728. <https://doi.org/10.1037/0022-0663.94.4.719>
- Ouellette, G., & Beers, A. (2010). A not-so-simple view of reading: How oral vocabulary and visual-word recognition complicate the story. *Reading and Writing*, 23(2), 189–208. <https://doi.org/10.1007/s11145-008-9159-1>
- 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(3), 554–566. <https://doi.org/10.1037/0022-0663.98.3.554>
- Salas, N., Llauradó, A., Castillo, C., Taulé, M., & Martí, M. A. (2016). Linguistic correlates of text quality from childhood to adulthood. In J. Perera, M. Aparici, E. Rosado, & N. Salas (Eds.), *Written and spoken language development across the lifespan: Essays in honour of Liliana Tolchinsky* (pp. 307–326). Springer.
- Salas, N., & Tolchinsky, L. (2017). Hunting for the links between word-level writing skills and text quality. In E. Segers & P. van den Broek, (Eds.), *Developmental perspectives in written language and literacy: In honor of Ludo Verhoeven* (pp. 103–118). John Benjamins Publishing Company.
- Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, 64(4), 913–951. <https://doi.org/10.1111/lang.12077>
- Schwartz, M., & Katzir, T. (2012). Depth of lexical knowledge among bilingual children: The impact of schooling. *Reading and Writing*, 25(8), 1947–1971. <https://doi.org/10.1007/s11145-011-9308-9>
- Sénéchal, M., Ouellette, G., & Rodney, D. (2006). The misunderstood giant: On the predictive role of early vocabulary to future reading. In D. K. Dickinson & S. B. Neuman, S. B. (Eds.), *Handbook of early literacy research* (Vol. 2, pp. 173–182). The Guildford Press.
- Tolchinsky, L. (2015). Continuidades y discontinuidades en el aprendizaje de la lengua escrita. In J. Mata, M. P. Núñez, & J. Rienda (Coords), *Didáctica de la lengua y la literatura* (pp. 131–146). Pirámide.
- Tolchinsky, L. (2019). Evolving structure of descriptive texts and learners' abilities. *Journal of Literacy Research*, 51, 293–314. <https://doi.org/10.1177/1086296X19858354>

- Verhallen, M., & Schoonen, R. (1998). Lexical knowledge in L1 and L2 of Third and Fifth Graders. *Applied Linguistics*, 19(4), 452–470. <https://doi.org/10.1093/applin/19.4.452>
- Verheyden, L., Van den Branden, K., Rijlaarsdam, G., Van den Bergh, H., & De Maeyer, S. (2010). Written narrations by 8-to 10-year-old Turkish pupils in Flemish primary education: A follow-up of seven text features. *Journal of Research in Reading*, 33(1), 20–38. <https://doi.org/10.1111/j.1467-9817.2009.01430.x>
- Wagner, R. K., Puranik, C. S., Foorman, B., Foster, E., Wilson, L. G., Tschinkel, E., & Kantor, P. T. (2011). Modeling the development of written language. *Reading and Writing*, 24(2), 203–220. <https://doi.org/10.1007/s11145-010-9266-7>