Read Writ (2016) 29:1317–1336
DOI 10.1007/s11145-016-9637-9

# Analysis of lexical quality and its relation to writing quality for 4th grade, primary school students in Chile

**Gabriela Gómez Vera**[1,3] · **Carmen Sotomayor**[1] · **Percy Bedwell**[2] · **Ana María Domínguez**[2] · **Elvira Jéldrez**[1]

Published online: 23 March 2016


**Abstract** Few studies have addressed vocabulary quality in developing writing skill in Spanish. Even less addressed it within the Chilean educational system. The specific objective of this study was to characterize, using a comprehensive set of indicators, the quality of the vocabulary produced by Chilean 4th grade students. Based on a national writing survey, a sample of 2056 texts written by 685 students was collected (narrative, persuasive, and informative texts). Current literature defines lexical quality as a composite of diverse factors that, while distinct, are interrelated. To represent the properties of the vocabulary, a set of indicators were selected: (a) lexical diversity; (b) lexical sophistication; and (c) lexical density. Using multilevel modeling (students and schools as levels 1 and 2) to explain a global writing score we found that diversity was a significant determinant for narrative and persuasive texts, density was a significant determinant for the three genres and sophistication was a significant determinant for narrative and expository text. In addition, indicators related to gender and socioeconomic conditions were only significant determinants of narrative stories. The parts of speech most often used also varied according to the purpose of each text. In all genres, words had a short extension and were very sensitive to the input presented in the stimuli. These results imply a significant challenge to this education system: how to promote the development of vocabulary in all children in order to support language learning.

**Electronic supplementary material** The online version of this article (doi:10.1007/s11145-016-9637-9) contains supplementary material, which is available to authorized users.

✉ Gabriela Gómez Vera
gabriela.gomez@ciae.uchile.cl

[1] Center for Advanced Research in Education, University of Chile, Santiago, Chile

[2] Arauco Fundation, Santiago, Chile

[3] Periodista José Carrasco Tapia N° 75, Santiago, Chile

**Keywords** Writing · Vocabulary · Lexical diversity · Lexical sophistication · Lexical density

## Introduction

Writing is a life skill applied in learning, communication, and social integration. However, the Chilean Education Quality Measurement System (SIMCE) has focused its studies exclusively on reading, and it was only in 2008 that the first pilot study was launched to measure writing quality among students (4th grade, 9–10 years old). The results were worrying, 38 % of evaluated students had only basic proficiency in writing (Unidad de Currículum y Evaluación, 2009). Despite these results, few investigations have addressed the possible underlying causes. Most research on the Chilean educational system focuses on general contextual determinants of achievement as the impact of socioeconomic conditions and school segregation (Contreras, Sepúlveda, & Bustos, 2010; Elacqua, 2012). Individual determinants taken into account are factors like gender and grade retention (Treviño et al., 2010), all of which play an important role in explaining educational achievement in the context of the Chilean educational system. Research on the role of writing components related to specific abilities such as vocabulary is less frequent in part due to the relevance of the contextual determinants and in part due to the major difficulty of conducting research on writing components at a national and systemic level. The current study tests how the understanding of vocabulary use and the understanding of writing achievements in the context of the Chilean educational system can benefit from a joint perspective.

The main focus is on lexical quality for two reasons, first, because it has been recently identified as a writing component in which Chilean student have low achievement rates (Agencia de Calidad de la Educación, 2013), consequently it is becoming a concern for policy and curriculum developers. Second, because from the point of view of writing acquisition, lexical quality is fundamental. Vocabulary is an ability that a writer needs to apply simultaneously with other abilities in order to produce a text (Smith-Lock, Nickels, & Mortensen, 2009). Among others, writing implies the use of vocabulary, orthography, and organizational structure, each of which corresponds to a cognitive task with specific challenges and specific developmental paths (Alamargot & Fayol, 2009). In addition, writing needs complex sub-processes such as idea generation and transcription to text (Flower & Hayes, 1981); these require vocabulary recognition, semantic and syntactic processing, and sentence and paragraph composition (Cuetos, 2009). Consequently, writing fluency is difficult to achieve for novice writers (Graham, Berninger, Abbott, Abbott, & Whitaker, 1997; Graham & Harris, 2000; McCutchen, 2011). From a cognitive point of view, the integration of these multiple tasks demands high on working memory, which has limited capacity (Sánchez, Moyano, & Borzone, 2011). A better mastery of aspects such as vocabulary knowledge and use is essential for the writer to be able to automatize the process and to decrease the cognitive load on working memory (McCutchen, 2011). Lexical quality is also related to reading development; previous studies have shown that a good mastery of vocabulary is positively related to

comprehension (Aarnoutse, van Leeuwe, & Verhoeven, 2005; Verhoeven, Van Leeuwe, & Vermeer, 2011), and it is precisely in the 4th grade when vocabulary is a particular determinant (Senechal, Ouellette, & Rodney, 2006) given that schoolwork no longer relies on simple decoding, but rather on understanding. Therefore, a limited vocabulary is a direct impediment to comprehension and learning.

Lexical quality, or richness (Ávila, 1991), is multidimensional, thus cannot be evaluated as a single indicator but rather through a composite of diverse factors that, while distinct, are interrelated (Malvern, Richards, Chipere, & Durán, 2004; Read, 2000). Concretely, the two principal concepts used when evaluating lexical quality are the quantity of vocabulary used in writing and the characteristics of the words used (McNamara, Crossley, & McCarthy, 2010; Perfetti & Hart, 2001). The concept of quantity refers to lexical diversity, or the number of unique words that the writer uses in the text (McNamara et al., 2010). The concept of quality, from the writer's point of view, denotes a broad diversity of vocabulary through the use of syntax and greater complexity. Two indicators related to these concepts are *lexical sophistication* and *lexical density*. Lexical sophistication refers to the employment of infrequent and more complex vocabulary and the use of longer words with more elaborate syllable structures. These words allow the writer "to express their meanings in a precise and sophisticated manner" (Read, 2000, p. 200). Lexical density refers to the quantity of content vocabulary present in a text. *Content Vocabulary* are words which have semantic relevance, both in isolation and within the context of the sentence; these words are normally identified as nouns, adjectives, non-auxiliary verbs, and adverbs. The other parts of speech, such as articles, pronouns, conjunctions, auxiliaries, etc., are categorized as functional words whose roles are linked to grammar and the production of formal text (Read, 2000). McNamara et al. (2010) have found that lexical diversity and word frequency are among the most predictive indicators of the quality of essays writing by college students. Olinghouse and Leaird (2009) found that diversity and the use of infrequent vocabulary where measures that increased in narrative texts produced by second and fourth grade students. In a comparison across genres, Olinghouse and Wilson (2013) observed variations in vocabulary use, informative texts having less diversity than narrative and persuasive texts and more content vocabulary use; this study also found that the vocabulary measures were related to writing quality but with variations across genres.

## Present study

The specific objective of the present study was to measure the lexical quality of 4th grade, primary school Chilean students through the use of a wide range of indicators. As noted by Ávila (1991) and Porras (2005), although this is an important factor within the framework of developing writing skills, there are few studies which analyze it in Spanish. Moreover, there is a lack of precise guidelines for teaching writing, and fewer have been established within the context of the Chilean educational system. A descriptive analysis of the vocabulary used by students represents a fundamental starting point for improving overall learning. In this context, an important contribution of the present study is distinguishing between

genres and to fit comparative models of writing achievement for narrative, persuasive and informative texts. A second contribution of the present study is a consideration for contextual aspects (Boscolo, 2008; Graham, Gillespie, & McKeown, 2013) such as gender and socioeconomic status. From this, it was possible to describe in detail the strengths and weaknesses of Chilean students in regard to vocabulary, while controlling for contextual determinants.

## Methods

### Sampling and coding

This study analyzed the lexical properties used by 4th grade students when writing texts for a nationwide evaluation developed by the Chilean Education Quality Measurement System (SIMCE; Unidad de Currículum y Evaluación, 2009). This exam was given to a representative sample of 4th grade students. In total, 22,854 students distributed across 706 educational establishments were evaluated, accounting for 9 % of total enrolled students at this grade level. Of this total, 687 students were selected for in-depth analyses in the present study, considering a total sampling error of 3.5 % and conserving the original group distribution in regard to gender, the geographic location of schools (North, Central, South, and Metropolitan Regions), the administrative dependence of schools (Municipal, Subsidized-Private, and Private), and socioeconomic status (Lower–Lower Middle, Middle, or Upper Middle–Upper). Each student wrote three pieces of text: (1) a narrative story for which the prompt was a photograph of a boy and dolphin in the water, accompanied by the instructions "Write a story based on this photo"; (2) a persuasive letter in which the prompt was "Your class collected money for a fieldtrip. They chose you to convince the school principal to give your class permission to go"; and (3) an informative news report in which the prompt was "It was learned yesterday that an investigator discovered dinosaur remains close to your school. Write a news report about this discovery". The resulting sample was a pool of 2056 texts. The writing samples were digitized following these criteria: (a) Orthographic errors were corrected in the transcription, both literal and in relation to accent marks; (b) Punctuation marks were not transcribed; (c) Hypo- and hyperpigmentation were corrected; (d) illegible words were marked as "NN"; (e) All written elements from the response sheet were transcribed, including elements external to text content (e.g. writing signature at the end of the persuasive letter). In particular cases requiring unforeseen changes, new corrections were made (see details in online supplementary material). The main goal was to generate a word database completely free of errors related to orthography, sentence structure or handwriting. The resulting database contained 134,168 words comprising 568,046 characters.

### Relationship between writing prompts and gender image

There is a key semantic aspect considering the explicitness of gender through morphemes in Spanish words. All of the prompts in this assessment made references

to objects and people in the masculine form—the photo provided for the narrative text was of a boy with a dolphin; the persuasive letter was to be addressed to the [male] principal; and the dinosaur bones by the school were discovered by a [male] investigator. While it could be argued that the prompts only served as a starting point for the students, a revision of the genders used in the writing samples highlighted the lack of female references (see detail, "Appendix 3"). The word *niño* (boy*)* was used 1441 times, while the word *niña* (girl) was used only 32 times in the texts, even though 49.5 % of the sample group was female. The word *investigador (male investigator)* was used 266 times compared to the use of *investigadora* (female investigator) only twice. Consequently, there was a tendency to promote masculine forms induced by the prompts. A marker of this tendency was observed through the use of the headword *teacher*, which was more used in its female forms: *profesora (*female teacher*)* was used 38 times and *profesor* (male teacher) was used 23 times. This demonstrates that when there was no restriction imposed by the prompts, the use of gender changed.

## Indicators

To represent the quantitative and qualitative aspects of vocabulary, several validated indicators were analyzed in this study, these are detailed next:

Lexical diversity: this measure refers to the range of vocabulary used in a text. The classical way to determine lexical diversity is by dividing the number of unique words by the total words in a text, which is referred to as the type/token ratio (TTR, McCarthy & Jarvis, 2010). However, this measurement is affected by the length of the text, where a higher quantity of words is associated with lesser word diversity (Malvern et al., 2004). To control for this, the present study calculated the corrected type/token ratio (CTTR). This corrected measurement divided the unique quantity of words by twice the square root of total words, instead of the absolute total words in a text (Malvern et al., 2004). However, Malvern et al. (2004) affirmed that a corrected measurement was still sensitive to an increased word count (Olinghouse & Leaird, 2009; Olinghouse & Wilson, 2013). Despite this, the impact of word count was only found significant for texts with hundreds of words (Malvern et al., 2004), a situation not applicable to the texts analyzed in the present study (see Table 1). More recent indicators exist that are not affected by text length, such as the Measure

**Table 1** Total and unique word counts, mean, and standard deviation by genre and in total

| Text | Number of texts | Mean of total words | SD of total words | Mean of varied words | SD of varied words | % of unique words in relation to total mean |
|---|---|---|---|---|---|---|
| Narrative | 681 | 99.965 | 33.832 | 58.253 | 16.277 | 60.17 % |
| Persuasive | 686 | 49.515 | 22.278 | 37.210 | 13.289 | 78.56 % |
| Informative | 686 | 46.796 | 23.368 | 33.456 | 13.580 | 75.38 % |
| Total | 2053 | 65.338 | 36.382 | 42.936 | 18.088 | 71.4 % |

of textual lexical diversity (MTLD, McCarthy & Jarvis, 2010) or the index D (Malvern et al., 2004). However, according to McNamara et al. (2010) the MTLD requires texts of <100 words, which is longer than the average length of the texts under current analysis. As reported in MacWhinney (2000) the index D requires an input of at least 50 different words in a text to be calculated, half of the sample do not achieve this minimum amount.[1]

Lexical sophistication: As previously mentioned, this notion alludes to the use of infrequent and more complex vocabulary that allows the writer greater precision in expression (Read, 2000). To quantify the sophistication of sampled texts, word length was taken as a reference, as measured through the number of characters in each word. Then, the proportion of polysyllabic words was calculated, defined as words with six or more characters, which was equal to the sum of the average and standard deviation (SD) of word length in the sample group. Following this, the proportion of polysyllabic words present in each text was used as an indicator in analyses. This measurement was employed as a reference given that the existing literature confirms that the use of longer words serves as a representation of lexical sophistication (Crossley, Weston, McLain Sullivan, & McNamara, 2011).

Lexical density: Each lexical item was classified in terms of parts of speech as nouns, adjectives, verbs, and adverbs (content words) or as pronouns, articles, conjunctions, and prepositions (functional words). Classification of parts of speech was performed using the TreeTagger software (Schmid, 1994). Furthermore, this program allowed for the separation of auxiliary verbs from functional verbs, in addition to identifying the headwords to which each word was associated. The headword is a guideword or dictionary reference that in texts can be identified through some of its inflections (López-Mezquita Molina, 2005). For example, when acting as an infinitive, the verb "to be" is a headword, whereas all of its conjugations are variations through which the headword is represented. In relation to the writers' vocabulary knowledge, headwords are a better reference than the effectively written words (Read, 2000). Using both criteria (headwords and lexical category), it was possible to differentiate frequency not in terms of total words but rather in relation to headwords, similar to the work carried by Porras (2005). This was the way in which the nouns, verbs, and adjectives used by students in each text were reported. Moreover, by applying the methodology used in calculating the CTTR, an analogous measurement was created that only considered the content headwords (CH), resulting in a CH-CTTR that provided an estimated lexical density for each essay.

## Data analyses

In addition to the indicators of lexical quality studied, other indicators of writing quality were also analyzed, specifically: adequacy to the communicative situation, coherence, cohesion, and overall text structure. These indicators were evaluated

[1] For those text in the sample with more than 50 different words, the index D was calculated using VOCD procedure of CLAN program (MacWhinney, 2000). The correlation between D index and CTTR measure was calculated for each genre: Narrative $r = 0.763$, N = 630; Persuasive $r = 0.670$, N = 296; Informative $r = 0.757$, N = 271.

using a rubric specifically developed for each dimension and for each type of text evaluated (Sotomayor et al., 2014). With the objective of obtaining a referential measurement regarding writing quality, a principal component analysis (PCA) was performed for the evaluated dimensions, from which specific indices with high internal consistency were developed for each text type ($\alpha > 0.79$). In each case, these indexes considered significantly correlated dimensions (for narrative texts: coherence, structure, cohesion and overall assessment; for persuasive texts: adequacy, structure, cohesion and overall assessment; for expositive texts: adequacy; coherence and overall assessment). The differentiations made between indicators according to genre were in agreement with findings by Olinghouse and Wilson (2013), who found variations in predictors of writing quality depending on text type. The indices were used as indicators of writing quality for each type of text; with all indices standardized on a scale T to obtain a median of 50 and a standard deviation of 10.

Correlations, linear regression and multilevel models were used to analyze the relationship of these indices with the indicators of lexical quality. In addition, the regressions controlled for reading achievement obtained in the same national assessment, student gender (coded as 1 for females and 0 for males) and socioeconomic status. This final factor was calculated through PCA considering parental education and household income ($\alpha = 0.79$), the variable had a median of 0 and a standard deviation of 1. The data for these two factors was obtained through a survey administered by SIMCE to students' parents for the corresponding year evaluated.

Survey answers and reading scores were affected by missing data, in order to avoid a reduction in the sample, a multiple imputation procedure was applied (Yuan, 2011) in the calculation of multilevel models. Since most of the variables were constructed specifically for the present research, the properties of each variable were analyzed. For the correlations (see Table 5), it was verified that no pair of variables were so strongly correlated that a possible collinearity could exist. The conventional limits of 80 or 90 % were not observed (Bressoux, 2008). Moreover, a multicollinearity analysis was made, as measured by calculating *tolerance* and the *inflation factor of variance* (IFV) for each variable (Foster, Barkus, & Yavorsky, 2006; Weisberg, 2005). It did not indicate collinearity (IFV $< 2.56$ in all cases). Altogether, this validated the consideration of each variable as an independent determinant in the constructed models. Once the independent determinants were estimated, a distribution analysis of the residuals showed that the regressions were statistically sufficient enough for inferences to be made (see descriptive statistics in "Appendix 2").

## Results

### Corpus characterization

The writing prompt was found to have a strong influence on the types of vocabulary used in the essays written by students. This result was both expected and positive in regards to the given task, with the vast majority of students adjusting to the themes

**Table 2** Indicator of lexical quality, means and differences among genres (repeated measures ANOVA)

| Indicators | Narrative | | Persuasive | | Informative | | Mauchly's test $\chi^2$ (2) | Greenhouse-Geisser $\varepsilon$ | Corrected $F$ test | Multiple comparisons |
|---|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | | | | |
| Diversity (CTTR) | 4.11 | 0.02 | 3.72 | 0.02 | 3.44 | 0.02 | 9.85* | 0.986 | 360.19* (1.97, 1338.70) | Nar > Per > Inf |
| Density (CH-CTTR) | 3.20 | 0.03 | 2.44 | 0.02 | 2.36 | 0.02 | 42.21* | 0.971 | 715.02* (1.94, 1318.61) | Nar > Per > Inf |
| Sophistication (% polysyll.) | 13.62 | 0.19 | 21.93 | 0.26 | 26.11 | 0.27 | 20.56* | 0.943 | 762.97* (1.89, 1280.69) | Inf > Per > Nar |

N = 680; *F* test, Greenhouse-Geisser corrections, degrees of freedom in parentheses

Multiple comparisons test with Šidàk corrections, $\alpha = 0.001$

*Nar* Narrative, *Per* Persuasive, *Inf* Informative

* $p < 0.001$

and genres given by the prompt. Moreover, the influence of the prompts were also observed in that the most frequent words were not used spontaneously by students, but rather as a direct response to the stimulus (see "Appendix 1"). For example, in writing the narrative story, a large majority of students used the word *delfín* (dolphin) or a derivative as well as the word *niño* (boy). The prompt for the persuasive letter could have also increased the use of longer words, since it was an informative news about a dinosaur discovery, the vast majority of students incorporated the words *school, investigator*, and *dinosaur* in their essay, and some, also, included the names of related professions such as *arqueólogo* (archaeologist) or *científico* (scientist), in addition to using dinosaur names such as *allosaurus, spinosaurus*, or *pterodactyl*.

Lexical diversity: As indicated in Table 1, students wrote texts containing 65 words on average. Of these, approximately 43 were unique words, while the rest were repetitions. When considering genre (see Table 1), a marked difference was found between narratives and the other text types. Specifically, text length of narrative stories was twice that of the persuasive letters and informative news reports.

Narratives presented the more varied word usage but also the lowest percentage of unique words. This shows the effect that text length can have on word diversity. That is to say, narrative, which were much longer than the other text types, exhibited a reduced diversity in vocabulary as the quantity of words increased.

Table 2 presents a comparison of means among genres for the diversity, density and sophistication measures. ANOVA with repeated measures with a Greenhouse-Geisser correction determined that the three indicators differed significantly between genres. Violations of the assumption of sphericity was found through Mauchly's Test in the case of diversity: $\chi^2$ (2) = 9.845, $p < 0.007$; density: $\chi^2$ (2) = 42.213, $p < 0.001$ and sophistication: $\chi^2$ (2) = 20.561, $p < 0.001$. Therefore F-statistics were calculated using Greenhouse-Geisser correction. Based on these estimations, a multiple comparisons test with Šidàk corrections ($\alpha = 0.001$) was applied (Abdi, 2007). Regarding lexical diversity, the results presented in Table 2 show that there were significant differences between averages according to text type, F (1.97, 1338.70) = 360.19, $p < 0.001$. The narrative stories presented the greatest diversity (M = 4.11), followed by persuasive letters (M = 3.72), and, finally, by informative news reports (M = 3.44). In relation to lexical density, the CH-CTTR showed significant differences between text genres, F (1.94, 1318.61) = 715.02, $p < 0.001$. There were greater density in narrative stories, followed by persuasive letters, and informative news reports. Regarding lexical sophistication there were significant differences in the use of polysyllable words between genres, F (1.89, 1280.69) = 762.97, $p < 0.001$. As expected, informative reports presented the higher proportion of polysyllable words (26.11) whereas narrative stories presented the lowest proportion (13.62).

The data showed that words used by 4th grade students were short, both when considering all texts and specific text types; the sampled texts contained approximately four characters per word and the average was 12 polysyllable words per text (see Table 3).

Lexical density: Regarding content and functional vocabulary, close to half of the words were content words, while the other half were functional.

In order to further understand the parts of speech and how these were distributed according to text type, an ANOVA for repeated measurements was used to test significant differences for each category. To correct the estimations, the same procedures previously described were applied. Table 4 presents the comparisons, showing that the two most used parts of speech were nouns and verbs, 21 % of words in narrative texts were nouns and 22 % were verbs; in persuasive texts 25 % of words were nouns and 23 % were verbs; in informative texts 27 % of words were nouns and 19 % were verbs. In addition, significant differences were observed in regard to the usage frequency for each part of speech, as dependent on genre and syntactic category. Regarding content vocabulary, for narrative stories, verbs were the most represented, followed by nouns and pronouns. This result is coherent with the narrative genre, where actions, textually represented by verbs, play a principal role, while the objects of these actions, textually represented by nouns and prenouns, would also have high usage. In the case of persuasive letters, the most used part of speech was nouns, followed by verbs, and, lastly, by pronouns. Moreover, verbs and pronouns were more used in the persuasive letters than in the other genre. This is also coherent with the genre, given that persuasive texts directly interact with the receptor, inviting them to respond through action. Finally, informative news reports presented the greatest incidence of nouns, followed by verbs, and, thirdly, by prepositions. This distribution can also be seen as pertinent to the purpose of the expository article required by the prompt.

Given that nouns and verbs were the most used types of speech, it is worthwhile to further explore their meanings. A list of the 50 most used nouns and verbs headwords is presented as supplementary material. For nouns, even when accounting for words directly taken from the prompt, the context and theme of the prompt had an influence. Likewise, words related to the sea, beach, and certain marine animals were very common and frequently mentioned in the narrative stories. Another two themes were the family and school. Regarding the former, words as *mother, father, son, uncle,* and *family* were found. Regarding the latter, words as *school, student,* and *classmate* were frequent. The headword "teacher" accounted for the Spanish gendered words *profesor* [male teacher] (38), *profesora* [female teacher] (90), *profesores* [plural male/mixed gender] (23), *profesoras* [plural-female] (6), and "*profe*" [neutral] (9).

**Table 3** Mean of characters and polysyllable words by text type and in total

| Text | Mean of characters per word | SD of characters per word | Mean of polysyllable words | SD of polysyllable words |
|---|---|---|---|---|
| Narrative | 3.892 | 2.222 | 13.543 | 6.330 |
| Persuasive | 4.366 | 2.431 | 10.671 | 5.164 |
| Informative | 4.824 | 3.010 | 11.845 | 5.787 |
| Total | 4.235 | 2.513 | 12.016 | 5.895 |

**Table 4** Parts of speech, mean frequency by text (%), and significant differences between text types

| Parts of speech | Narrative | | Persuasive | | Informative | | Mauchly's test $\chi^2$ (2) | Greenhouse-Geisser $\varepsilon$ | Corrected $F$ test | Multiple comparisons |
|---|---|---|---|---|---|---|---|---|---|---|
| | *M* % | *SD* | *M* % | *SD* | M % | SD | | | | |
| Nouns | 20.61 % | 0.19 | 25.45 % | 0.28 | 27.21 % | 0.30 | 27.58*** | 0.96 | 212.98*** (1.92, 1305.95) | Inf > Per > Nar |
| Verbs | 22.27 % | 0.15 | 22.78 % | 0.22 | 19.49 % | 0.23 | 39.37*** | 0.95 | 81.01*** (1.89, 1285.48) | (Nar = Per) > Inf |
| Adjectives | 6.14 % | 0.12 | 5.15 % | 0.15 | 6.65 % | 0.17 | 45.00*** | 0.94 | 28.46*** (1.88, 1276.05) | (Nar = Inf) > Per |
| Adverbs | 4.52 % | 0.08 | 3.13 % | 0.11 | 4.96 % | 0.15 | 50.32*** | 0.93 | 65.16*** (1.87, 1267.35) | (Nar = Inf) > Per |
| Pronouns | 9.56 % | 0.14 | 12.02 % | 0.18 | 5.95 % | 0.17 | 42.21*** | – | 354.97*** (2, 1358)[a] | Per > Nar > Inf |
| Articles | 14.38 % | 0.15 | 6.27 % | 0.16 | 12.45 % | 0.21 | 15.25*** | 0.98 | 645.27*** (1.96, 1328.45) | Nar > Inf > Per |
| Conjunctions | 11.13 % | 0.14 | 11.15 % | 0.17 | 7.61 % | 0.19 | 30.66*** | 0.96 | 182.87*** (1.92, 1300.50) | (Nar = Per) > Inf |
| Prepositions | 8.33 % | 0.11 | 10.38 % | 0.17 | 13.03 % | 0.19 | 34.50*** | 0.95 | 222.29*** (1.91, 1293.82) | Inf > Per > Nar |
| Other | 3.05 % | 0.08 | 3.64 % | 0.12 | 2.62 % | 0.11 | 30.89*** | 0.96 | 24.60*** (1.91, 1300.11) | Per > (Nar > Inf [b]) |

N = 680; *F* test, Greenhouse-Geisser corrections, degrees of freedom in parentheses

Multiple comparisons test with Šidàk corrections, $\alpha = 0.001$

*Nar* Narrative, *Per* Persuasive, *Inf* Informative

*** $p < 0.001$

[a] Sphericity assumed

[b] Significant at 0.003

Verb frequencies included copulative and auxiliary verbs, with the resulting observations finding that the verbs *ser* and *estar* (English *to be*, represented by two distinct verb forms in Spanish) were those most used by students. It is also worth mentioning that the verb *encontrar* (to find) was repeatedly used—even though this was not presented in the narrative prompt, it is closely related to the concept of discovery, which was used in the prompt. It is possible that students resorted to *encontrar* as an alternative for *descubrir* (to discover), revealing a strategy for searching synonyms.

The frequency distribution was similar between nouns and verbs. Moreover, the general tendency in vocabulary selection was that a reduced number of words were commonly used in texts whereas many words were used very little or only once. In addition to this, even headwords found at the most-used 50 headwords list had quite low appearance frequencies, the nouns *miedo* (fear) and *barco* (ship) used 60 times; the verbs *seguir* (to follow) and *pescar* (to fish) used 73 and 68 times, respectively.

Table 5 shows the quantity of words used between only 1 and 5 times in all texts. While the proportion that these words represent is small (4.12 %), this share is significant when considering the total number of words, which surpasses 134,000, that means that each word represent a 0.0007 % of the total.

## Interpretative analyses

Correlation analyses: Table 6 presents the results of the three indicators analyzed for each text type. In all cases, the most important correlation was found between diversity and density ($r = 0.75$ in narratives and persuasive texts, $r = 0.77$ in informative texts). The direction of the correlation between sophistication (represented by the proportion of polysyllabic words) and the other two indicators vary across genre. For narrative stories, sophistication was positively correlated to diversity ($r = 0.21$) and to density ($r = 0.11$). For persuasive letters, sophistication was correlated only to diversity in a low magnitude ($r = 0.08$). For informative news, sophistication was correlated only to diversity and this correlation was negative ($r = -0.15$). The results confirm that connections between indicators vary across genres. Consequently, a regression model was fitted for quantifying the role that each determinant played while taking covariance into consideration.

Multilevel regressions: This model allowed estimating the influence that each indicator had while also taking into account possible interactions and the hierarchical structure of the educational system (Hox, 2010). The regression procedure implies that writing quality could be determined by lexical quality; with lexical quality represented by the three previously described indicators and writing quality represented by the grade scores given to each student in the general evaluation of each text. Table 7 present 3 models, each one fitted to explain the writing achievement of a genre. As previously stated, gender, socioeconomic conditions and school segregation are known for having a great impact in learning outcomes in the Chilean educational system. The first two aspects are controlled by adding variables to the model, and the third by its multilevel structure that considers schools as a source of variance.

The explanatory capacity of each model shown in Table 7 is represented by the proportion of variance explained, which uses the null model as reference (see

"Appendix 4"). At the school level, the most effective model is the one fit to explain narrative outcomes, this model accounted for 81.19 % of the score variance from school to school. In the case of variance within the school, the model fit to explain the score of persuasive letters is more effective (32.03 %). The variance explained by the three models is important but not different from that estimated in previous studies about this educational system, the main difference is that those studies used reading and mathematics scores as outcome (Valenzuela et al., 2015).

In addition to differences in the explicative capacity of the models altogether, there were also differences found in the estimated parameters of each factor. The estimated parameter of diversity is larger in the case of persuasive texts (4.16), smaller in the case of narrations (2.76) and it is not significantly different from zero in the case of informative news. The estimated parameters for density is larger in the case of narrative texts (4.39) and smaller in the case of informative news (2.05). The estimated parameters of sophistication are equally significant in the case of narrative and persuasive texts but with a different direction, positive in the case of persuasive letters (0.17) and negative in the case of narrations (-0.15). These results confirmed that indicators of vocabulary quality were determinants of the writing outcomes but with variations across genres.

Regarding control factors, there were variations across genres too. SIMCE reading score was a significant determinant in all cases. The parameter estimated for the models of informative text is particularly significant, considering that the scale used by SIMCE was set for an average of 250 points. Accordingly, a student that obtained an average score in reading is expected to obtain more than 22 points in writing ($250 \times 0.09 = 22.50$) which is twice the standard deviation. In the case of narrative texts, the parameter estimating for reading is smaller (0.05) but also significant. Socioeconomic status is only significant in the context of narrative texts (0.93). In addition, the estimated parameters for SES are smaller than those estimated for the lexical determinants. Finally, according to the models, females showed a small but positive effect only in the case of narrative texts (1.61). According to the models, the results of persuasive letters and informative news reports did not vary across gender.

## Discussion

The objective of this study was to characterize, using a comprehensive set of indicators, the quality of the vocabulary produced by Chilean 4th grade students. The analysis showed that the words had a short extension and were sensitive to

**Table 5** Frequency of least used words

| Words | *N* | % |
|---|---|---|
| Words used five times | 236 | 0.18 |
| Words used four times | 334 | 0.25 |
| Words used three times | 580 | 0.43 |
| Words used two times | 1080 | 0.81 |
| Words used only once | 3293 | 2.45 |
| Total number of words used from 1 to 5 times | 5523 | 4.12 |

Total words = 134,138

**Table 6** Correlation between indicators of lexical quality according to text type

| Narrative | | Persuasive | | Informative | |
|---|---|---|---|---|---|
| Density | Sophistication | Density | Sophistication | Density | Sophistication |
| Diversity | | | | | |
| 0.75*** | 0.21*** | 0.75*** | 0.08* | 0.77*** | −0.15*** |
| Density | | | | | |
| – | 0.11** | – | 0.07 | – | 0.00 |

N narrative = 681; N Persuasive 686; N Informative = 686

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

**Table 7** Multilevel linear regression estimations by text type

| | Narrative | Persuasive | Informative |
|---|---|---|---|
| Fixed effects | | | |
| Diversity | 2.76*** | 4.16*** | 0.45 |
| | (0.80) | (0.77) | (0.84) |
| Density | 4.39*** | 3.28*** | 2.05* |
| | (0.68) | (0.84) | (0.92) |
| Sophistication | −0.15** | 0.17** | 0.07 |
| | (0.06) | (0.04) | (0.05) |
| Reading achievement | 0.05*** | 0.05*** | 0.09*** |
| | (0.01) | (0.01) | (0.01) |
| Gender | 1.61** | 1.02 | 0.28 |
| | (0.59) | (0.61) | (0.67) |
| Socioeconomic status | 0.93** | 0.55 | 0.39 |
| | (0.35) | (0.35) | (0.39) |
| Intercept | 11.91*** | 8.07*** | 19.01*** |
| | (2.33) | (2.24) | (2.45) |
| Random effects | | | |
| Intra-class correlation | 0.05 | 0.07 | 0.05 |
| Level 2 (school) variances | 3.08 | 4.46 | 3.47 |
| Variance explained | 81.19 % | 72.38 % | 78.51 % |
| Level 1 (students) variance | 55.62 | 55.77 | 67.72 |
| Variance explained | 31.92 % | 32.03 % | 19.17 % |
| −2 log V | 4718.22 | 4753.02 | 4870.92 |
| Δ | 319.42 | 308.38 | 203.28 |
| N students | 683 | 686 | 686 |
| N schools | 272 | 272 | 272 |

*S*tandard errors in parentheses

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$l

the input presented in the stimuli. The most used nouns and verbs were few and reduced to the simplest forms. However, in the case of Chile, data does not exist for comparing the present observations with the vocabulary obtained by students during other stages of vocabulary development; due to which it is impossible to confirm whether the low level of lexical development in 4th grade students is a permanent condition throughout schooling or if it is an impermanence related to the early learning stage that can be compensated during later education.

To represent the properties of the vocabulary, a set of indicators were selected: (a) lexical diversity; (b) lexical sophistication; and (c) lexical density. Following the study of Olinghouse and Wilson (2013) the interpretative models showed that these indicators of lexical quality were related to a general measure of writing quality presenting variations across genres. The models allowed to observe that diversity was a significant determinant for narrative and persuasive texts, density was a significant determinant for the three genres and sophistication was a significant determinant for narrative and expository text. These results are in agreement with those obtained by Olinghouse and Wilson (2013), who also found that the predictive capacity of lexical determinants in regards to writing varied according to genre. These authors concluded also that lexical diversity was the principal lexical determinant in writing quality. In the present study, this role was shared by both lexical diversity and density and varied across genres. On the other hand, Olinghouse and Wilson found that diversity was similar between narrative and argumentative texts, a result not found in the present analyses. In order to explain the difference in these results, a more comparable tasks and procedures would be necessaries.

The incorporation of the reading comprehension score in the models allows to conclude that even contrasting with this more general achievement indicator, lexical determinants have a significant effect over writing achievement. Regarding socioeconomic status and gender, the estimations are only significant determinants of the writing outcomes of narrative texts. This result is similar to the finding of Troia, Harbaugh, Shankland, Wolbers, and Lawrence, (2013), who found that female students wrote better fiction than male students. This implies a need of balancing gender mentions when designing the tests used for writing evaluations. The two aspects are generally viewed as common determinants of learning achievement (Logan & Johnston, 2010; Ma, 2008; Elacqua, 2012). The present results suggest that the effect of socioeconomic status and gender on writing could vary according to gender. Consequently in educational evaluation, there could be variations in achievement if writing is measured using different text types.

The vocabulary used by students in their writing was highly determined by the vocabulary used in the prompts. From a general learning point of view, such as for creativity and idea development, this dependence on the prompt can be seen as a limitation. However, from the point of view of lexical development, it is interesting to note the pedagogic opportunity represented by these results. Prompts that promote the use of more diverse and sophisticated vocabulary could be a suitable method for improving lexical quality, thereby increasing the vocabulary available to students and familiarizing them with a wider and more complex use of

the words. This finding is relevant from a pedagogic perspective and indicates the necessity of taking into account lexical quality when formulating lesson plans for writing, both at the curricular level as in the daily work of teachers.

**Acknowledgments** Funding from PIA-CONICYT Basal Funds for Centers of Excellence Project BF0003 and CONICYT Project of Insertion of Advanced Human Capital in the Academy No. 79112008 are gratefully acknowledged.

## Appendix 1: Usage frequency of headwords suggested through the writing prompt

Nouns suggested through the writing prompts

| Nouns | Frequency |
|---|---|
| *Delfín* (Dolphin) | 2505 |
| *Niño* (Boy) | 1691 |
| *Dinosaurio* (Dinosaur) | 1291 |
| *Paseo* (Walk) | 962 |
| *Hueso* (Bone) | 959 |
| *Director* (Director) | 842 |
| *Permiso* (Permission) | 588 |
| *Curso* (Course) | 519 |
| *Resto* (Remain) | 428 |
| *Escuela* (School) | 411 |
| *Investigador* (Researcher) | 266 |
| *Plata* (Silver) | 188 |
| *Carta* (Letter) | 185 |
| *Descubrimiento* (Discovery) | 112 |
| *Noticia* (News) | 122 |

Verbs suggested through the writing prompts

| Verbs | Frequency |
|---|---|
| *Ir* (To go) | 2188 |
| *Dar* (To give) | 703 |
| *Descubrir* (To discover) | 255 |
| *Escribir* (To write) | 76 |

## Appendix 2: Descriptive statistics for variables in mixed models

Narrative stories (N = 683)

| Variable | M | SD |
|---|---|---|
| CTTR index (Variety) | 4.11 | 0.58 |
| LC-CTTR index (Density) | 3.20 | 0.66 |
| Polysyllable frequency (%) | 13.62 | 4.98 |
| Reading score | 264.51 | 52.21 |
| Gender | 0.50 | 0.50 |
| Socioeconomic status | 0.00 | 1.00 |
| Narrative writing score | 50.10 | 9.86 |

Persuasive letters (N = 686)

| Variable | M | SD |
|---|---|---|
| CTTR index (Variety) | 3.72 | 0.61 |
| LC-CTTR index (Density) | 2.44 | 0.54 |
| Polysyllable frequency (%) | 21.89 | 6.77 |
| Reading score | 263.37 | 52.13 |
| Gender | 0.50 | 0.50 |
| Socioeconomic status | 0.00 | 1.00 |
| Argumentative writing score | 50.00 | 10.00 |

Informative news report (N = 686)

| Variable | M | SD |
|---|---|---|
| CTTR index (Variety) | 3.72 | 0.61 |
| LC-CTTR index (Density) | 2.44 | 0.54 |
| Polysyllable frequency (%) | 21.89 | 6.77 |
| Reading score | 263.20 | 52.13 |
| Gender | 0.50 | 0.50 |
| Socioeconomic status | 0.00 | 1.00 |
| Expository writing score | 50.02 | 9.99 |

## Appendix 3: Variations between genders in the use of words suggested through the writing prompts

Gender and number variation of the headword *niño* (child)

| Word | Frequency |
|---|---|
| *Niño* (boy) | 1441 |
| *Niños* (boys) | 180 |
| *Niña* (girl) | 32 |
| *Niñito* (little boy) | 25 |
| *Niñas* (girls) | 11 |
| *Niñita* (little girl) | 1 |

Gender and number variation of the headword *investigador* (researcher)

| Word | Frequency |
|---|---|
| *Investigador* [male researcher] | 266 |
| *Investigadores* [male researchers] | 51 |
| *Investigadora* [female researcher] | 2 |
| *Investigadoras* [female researchers] | 1 |

Gender and number variation of the headword *director* (principal)

| Word | Frequency |
|---|---|
| *Director* [male principal] | 842 |
| *Directores* [male principals] | 1 |
| *Directora* [female principal] | 56 |
| *Directoras* [female principals] | 0 |

## Appendix 4: Multilevel linear regression, null models (mean of iterations estimations regressions)

| | Narrative | Persuasive | Informative |
|---|---|---|---|
| Intercept (Thresholds) | 49.603*** (0.47) | 49.35*** (0.47) | 49.64*** (0.47) |
| Random effects | | | |
| Intra-class correlation | 0.17 | 0.16 | 0.16 |
| Level 2 (school) variance | 16.37 | 16.16 | 16.15 |
| Level 1 (student) variance | 81.69 | 82.04 | 83.78 |
| −2 log V | 5037.64 | 5061.40 | 5074.20 |
| N Students | 683 | 686 | 686 |
| N Schools | 272 | 272 | 272 |

Standard errors in parentheses

*** $p < 0.001$

## References


Aarnoutse, C., van Leeuwe, J., & Verhoeven, L. (2005). Early literacy from a longitudinal perspective. *Educational Research and Evaluation, 11*(3), 253–275.

Abdi, H. (2007). The Bonferronni and Šidák corrections for multiple comparisons. In N. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 103–107). Thousand Oaks, (CA): Sage.

Agencia de Calidad de la Educación. (2013). *Resultados Nacionales SIMCE Escritura 6° Básico 2012*. Santiago de Chile: MINEDUC.

Alamargot, D., & Fayol, M. (2009). Modelling the development of written composition. In R. Beard, D. Myhill, J. Riley, & M. Nystrand (Eds.), *The SAGE handbook of writing development* (pp. 23–47). Thousand Oaks, (CA): Sage.

Ávila, R. (1991). Densidad léxica y adquisición del vocabulario: niños y adultos. In *El español de América: actas del III congreso internacional del español en América: Valladolid, 3 a 9 de julio de 1989*, (pp. 621–630). Valladolid: Consejería de Cultura y Turismo.

Boscolo, P. (2008). Writing in primary school. In C. Bazerman (Ed.), *Handbook of research on writing: History, society, school, individual, text* (pp. 359–379). New York, (NY): Taylor & Francis.

Bressoux, P. (2008). *Modélisation statistique appliquée aux sciences sociales*. Bruxelles: De Boeck.

Contreras, D., Sepúlveda, P., & Bustos, S. (2010). When schools are the ones that choose: The effects of screening in Chile. *Social Science Quarterly, 91*(5), 1349–1368. doi:10.1111/j.1540-6237.2010.00735.x.

Crossley, S. A., Weston, J. L., McLain Sullivan, S. T., & McNamara, D. S. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication, 28*(3), 282–311. doi:10.1177/0741088311410188.

Cuetos, F. (2009). *Psicología de la escritura*. España: Wolters Kluwer Educación.

Elacqua, G. (2012). The impact of school choice and public policy on segregation: Evidence from Chile. *International Journal of Educational Development, 32*(3), 444–453. doi:10.1016/j.ijedudev.2011.08.003.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*(4), 365–387. doi:10.2307/356600.

Foster, J. J., Barkus, E., & Yavorsky, C. (2006). *Understanding and using advanced statistics: A practical guide for students*. London: Sage.

Graham, S., Berninger, V., Abbott, R., Abbott, S., & Whitaker, D. (1997). The role of mechanics in composing of elementary school students: A new methodological approach. *Journal of Educational Psychology, 89*(1), 170–182. doi:10.1037/0022-0663.89.1.170.

Graham, S., Gillespie, A., & McKeown, D. (2013). Writing: Importance, development, and instruction. *Reading and Writing, 26*(1), 1–15. doi:10.1007/s11145-012-9395-2.

Graham, S., & Harris, K. (2000). The role of self-regulation and transcription skills in writing and writing development. *Educational Psychologist, 35*(1), 3–12. doi:10.1207/S15326985EP3501_2.

Hox, J. (2010). *Multilevel analysis: Techniques and applications*. Great Britain: Routledge.

Logan, S., & Johnston, R. (2010). Investigating gender differences in reading. *Educational Review, 62*(2), 175–187. doi:10.1080/00131911003637006.

López-Mezquita Molina, M. T. (2005). *La evaluación de la competencia léxica: tests de vocabulario. Su fiabilidad y validez*. Ph.D. thesis, Universidad de Granada, Facultad de Filosofía y Letras, Departamento de Filología Inglesa.

Ma, X. (2008). Within-school gender gaps in reading, mathematics, and science literacy. *Comparative Education Review, 52*(3), 437–460. doi:10.1086/588762.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Basingstoke: Palgrave Macmillan.

McCarthy, P., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods, 42*(2), 381–392. doi:10.3758/BRM.42.2.381.

McCutchen, D. (2011). From novice to expert: Implications of language skills and writing-relevant knowledge for memory during the development of writing skill. *Journal of Writing Research, 3*(1), 51–68. doi:10.17239/jowr-2011.03.01.3.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication, 27*, 57–86. doi:10.1177/0741088309351547.

Olinghouse, N., & 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.

Olinghouse, N., & Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing, 26*(1), 45–65. doi:10.1007/s11145-012-9392-5.

Perfetti, C. A., & Hart, L. (2001). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–214). Amsterdam, the Netherlands: John Benjamins.

Porras, M. R. (2005). Una responsabilidad escolar olvidada: el desarrollo del componente léxico. *Revista Educación, 29*, 31–44.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Sánchez, V., Moyano, V., & Borzone, A. (2011). Demandas cognitivas de la escritura: comparación de dos situaciones de producción. *Estudios pedagógicos, 37*, 227–236.

Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of international conference on new methods in language processing*, (pp. 44–49). Manchester, UK.

Senechal, 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 (Eds.), *Handbook of early literacy research* (Vol. 2, pp. 173–182). New York, NY: Guilford.

Smith-Lock, K. M., Nickels, L., & Mortensen, L. (2009). Story writing skills of adults with a history language-impairment. *Reading and Writing, 22*(6), 713–734. doi:10.1007/s11145-008-9138-6.

Sotomayor, C., Gómez, G., Jeldrez, E., Bedwell, P., & Domínguez, A. (2014). Calidad de la Escritura en la Educación Básica. (Working Paper No 13). Santiago de Chile: CIAE. Retrieved from http://www.ciae.uchile.cl/download.php?file=doctrabajo/13-072014.pdf.

Treviño, E., Valdés, H., Castro, M., Costilla, R., Pardo, C., & Rivas, F. D. (2010). *Factores asociados al logro cognitivo de los estudiantes de América Latina y el Caribe*. Santiago de Chile: UNESCO – LLECE.

Troia, G., Harbaugh, A., Shankland, R., Wolbers, K., & Lawrence, A. (2013). Relationships between writing motivation, writing activity, and writing performance: Effects of grade, sex, and ability. *Reading and Writing, 26*(1), 17–44. doi:10.1007/s11145-012-9379-2.

Unidad de Currículum y Evaluación. (2009). *Informe de Resultados de Escritura SIMCE 2008*. Santiago de Chile: Ministerio de Educación.

Valenzuela, J., Gómez, G., & Sotomayor, C. (2015). The role of reading engagement in improving national achievement: An analysis of Chile's 2000–2009 PISA results. *International Journal of Educational Development, 40*, 28–39. doi:10.1016/j.ijedudev.2014.11.011.

Verhoeven, L., van Leeuwe, J., & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading, 15*(1), 8–25. doi:10.1080/10888438.2011.536125.

Weisberg, S. (2005). *Applied linear regression*. New Jersey: Wiley.

Yuan, Y. (2011). Multiple imputation using SAS software. *Journal of Statistical Software, 45*(6), 1–25.