


Assessing academic writing in Primary, Secondary, and Higher Education

Rocío Cuberos-Vicente 
Universidad Complutense de Madrid, Spain
Correspondence's author: marcuber@ucm.es

Verónica Martínez 
Universidad de Oviedo, Spain
martinezveronica@uniovi.es

Elisa Rosado 
Universidad de Barcelona, Spain
erosado@ub.edu

Melina Aparici 
Universidad Autónoma de Barcelona, Spain
melina.aparici@uab.cat

Received: 14-January-2025

Accepted: 24-July-2025

Abstract

Mastering the linguistic skills required for academic writing becomes a challenge at all educational stages. The present study analyses the relationship between the use of linguistic resources and the perceived quality of academic texts with the aim of developing valid instruments for the assessment of text quality. A total of 212 argumentative-expository texts from 65 primary school pupils, 78 secondary school pupils, and 69 university students were analysed. In a first analysis, the morphosyntactic, lexical, and discursive resources of texts were coded. In a second phase, teachers at each educational level evaluated them with an analytical rubric designed to assess linguistic and rhetorical aspects. The results showed a developmental increase in most of the linguistic features analysed, except for the proportion of discourse markers and lexical density. External evaluations showed greater variability in primary and secondary education than at university level, although scores did increase with educational level. The perceived quality of the texts was associated with different features at each level: while productivity was strongly linked to higher scores in primary and secondary education, in higher education a negative correlation emerged between the use of discourse markers and teacher evaluations. These findings will inform the design of evaluation instruments tailored to the specific requirements of each educational stage.

Keywords: Academic language; writing evaluation; writing achievement; Primary Education; Secondary Education; Higher Education.

How to cite: Cuberos-Vicente, R., Rosado, E., Martínez, V., & Aparici, M. (2026). La evaluación de la escritura académica en Educación Primaria, Secundaria y Universitaria. *Ocnos*, 25(1). https://doi.org/10.18239/ocnos_2026.25.1.588



La evaluación de la escritura académica en Educación Primaria, Secundaria y Educación Superior

Rocío Cuberos-Vicente 
Universidad Complutense de Madrid, España
Autor de correspondencia: marcuber@ucm.es

Verónica Martínez 
Universidad de Oviedo, España
martinezveronica@uniovi.es

Elisa Rosado 
Universidad de Barcelona, España
erosado@ub.edu

Melina Aparici 
Universidad Autónoma de Barcelona, España
melina.aparici@uab.cat

Recibido: 14-Enero-2025

Aceptado: 24-Julio-2025

Resumen

Dominar las habilidades lingüísticas requeridas en escritura académica es un desafío en todas las etapas educativas. Este estudio analiza la relación entre el uso de recursos lingüísticos y la calidad percibida en textos académicos con el objetivo de desarrollar instrumentos válidos de evaluación de la calidad textual. Se analizaron 212 textos analíticos de 65 alumnos de Educación Primaria, 78 de Secundaria y 69 universitarios. En un primer análisis, se codificaron recursos morfosintácticos, léxicos y discursivos de estos textos. Seguidamente, profesores de cada nivel educativo los evaluaron con una rúbrica analítica sobre aspectos lingüísticos y retóricos. Los resultados mostraron un aumento evolutivo en la mayoría de las características lingüísticas, excepto en la proporción de marcadores discursivos y la densidad léxica. La evaluación externa fue más diversa en Primaria y Secundaria que en la universidad, aunque las puntuaciones mejoraron con el nivel educativo. La calidad percibida de los textos se relacionó con características específicas en cada nivel: la productividad era clave para obtener buenas puntuaciones en Primaria y Secundaria, mientras que en Educación Superior se observó una correlación negativa entre el uso de marcadores del discurso y la evaluación docente. Estos hallazgos contribuirán al diseño de herramientas de evaluación adaptadas a cada etapa educativa.

Palabras clave: Lenguaje académico; evaluación de la escritura; desarrollo de la escritura; Educación Primaria; Educación Secundaria; Educación Superior.

Cómo citar: Cuberos-Vicente, R., Rosado, E., Martínez, V., & Aparici, M. (2026). La evaluación de la escritura académica en Educación Primaria, Secundaria y Universitaria. *Ocnos*, 25(1). https://doi.org/10.18239/ocnos_2026.25.1.588



INTRODUCTION

Reading a text almost inevitably involves evaluating its quality. As readers, our perception is shaped by factors ranging from the interest of the topic to the information provided or, crucially, the way in which this information is presented and structured. In general terms, evaluating a text is a fairly intuitive process. However, from the specialised perspective of researchers or teachers, text quality is reflected both in more or less subjective overall assessments, and in the observation of specific analytical criteria (Berman and Nir-Sagiv, 2010; Cuberos-Vicente, 2019). Furthermore, because the quality of a text is closely tied to its communicative effectiveness and to genre-specific expectations, it can partly be explained by examining its linguistic characteristics (McNamara et al., 2010). This study examines the alignment between researcher-based indicators of text quality and teacher-based assessments in texts produced by primary, secondary and university students.

In academic contexts, mastering the language skills required to produce high-quality texts is a challenge at all educational stages. Difficulties in the production of academic writing can lead to issues related to school, social and/or professional integration (Uccelli, 2023). Entering academic life entails becoming familiar with the conventions and language characteristics of texts, predominantly expository and/or argumentative, that are produced in academic settings (Schleppegrell, 2004; Snow and Uccelli, 2009). Although these genres are introduced in primary education and receive explicit attention in compulsory and post-compulsory secondary education, students continue to struggle with the appropriate use of linguistic resources and discursive strategies specific to academic texts when they transition to university. This remains a concern for teachers across levels and subjects, as difficulties in writing academic texts persist even beyond undergraduate level (Gavari-Starkie and Tenca-Sidotti, 2017; Rosado et al., 2021). Developing reliable tools to assess such texts is therefore essential for supporting children, adolescents, and adults in learning the features of academic writing. This study examines how specific linguistic characteristics influence readers' assessments of text quality. It focuses on analytical texts, those focused on analysis and argumentation (Schleppegrell, 2004; Uccelli et al., 2019) which typically combine both expository and argumentative components (Tolchinsky et al., 2017), and are commonly produced in academic settings. Compared with cognitively less demanding discourse genres (e.g., narratives), analytical texts require writers to articulate their point of view on a topic, present arguments for and against it, and offer appropriate evidence and/or possible refutations (Toulmin, 2002). Such demands make analytical texts particularly challenging for writers across ages/educational levels (Tolchinsky et al., 2017). While the organisation of written narratives tends to be mastered by the age of 10, analytical writing requires an advanced use of grammatical, lexical, and discursive forms and functions, which are acquired later in life (Uccelli et al., 2012). In these texts, authors are expected to use diverse yet precise vocabulary, to condense information through complex syntax, and organise discourse using intra- and inter-sentential connectivity devices that signal both textual transitions and author's stance (Aparici et al., 2021; Schleppegrell, 2004; Snow and Uccelli, 2009).

Previous studies have identified which lexical, morphosyntactic, and discourse resources serve as indicators of text quality and have examined how these indicators related to one another throughout development (Aparici et al., 2021; Crossley et al., 2011; Salas et al., 2016; Tolchinsky et al., 2021). However, the extent to which these elements influence text evaluation, i.e., how the use of specific features affects external judges' perceptions of 'quality' has not been examined from a developmental perspective.

An initial exploration of the relationship between linguistic indicators of text quality and external evaluations of this corpus was conducted in Tolchinsky et al. (2023) which analysed how aspects directly related to the content and structure of the texts helped explain perceived quality as judged by expert raters. Aparici et al. (2024) also explored which textual features predict teachers' evaluations, noting that they vary by educational level and/or exposure to specific pedagogical work. However, these studies only focused on the relationship between linguistic indicators and holistic evaluations.

OBJECTIVES

This study analyses the quality of analytical texts produced by students in primary, secondary and higher education, with the specific aim of:

1. Examining potential variation in the external evaluation of text quality across educational levels.
2. Analysing the development of a set of linguistic resources identified as indicators of text quality, comparing their use in texts across educational levels.
3. Determining which of these linguistic resources best account for external assessments of text quality at each educational level.

METHODS

Participants

A total of 212 students from three educational levels participated in this study: 6th grade primary school (n=65), 4th grade secondary school (n=78), and 1st and 2nd year university students (n=69). The texts were collected in León and Ciudad Real, in two primary schools, three secondary schools, and at the University of León. [Table 1](#) shows the mean, standard deviation, age range, and sex distribution of the participants.

Table 1

Average age, age range and sex of participants by educational level

Educational level	M Age (SD)	Age range	Participants		
			Male	Female	Total
Primary education	11.6 (0.28)	11.1–12.1	30	35	65
Secondary education	15.8 (0.67)	15.0–18.3	33	45	78
Higher education	21 (2.31)	19.10–31.8	27	42	69

Procedure

Writing task

During a teaching sequence designed to improve skills related to the production of analytical texts, participants wrote five texts on a computer, each within a maximum of 30 minutes, and no word limit. The teaching sequence was implemented over seven sessions in the regular classroom by the students' regular teachers, who had received specific training beforehand. The pedagogical activities included readings on the topic of the sequence—freedom of movement between countries—, class discussions on the characteristics of analytical texts, and peer-assessment tasks (see [Tolchinsky et al., 2023](#)). This study analyses text 4, which was produced immediately after the pedagogical activities. Text 4 was selected because participants had already read and written about this topic during the teaching sequence, ensuring that the texts were produced with identical instructions and within a controlled pedagogical context ([Tolchinsky et al., 2021](#)).

External evaluation of text quality

Nine experienced teachers from each educational level participated in the evaluation. They were provided with an analytical rubric comprising six assessment criteria: content, organisation, cohesion, vocabulary, correctness, and communicative impact. This rubric was developed as part of the *Analytical Writing and Linguistic Diversity* project (EDU2015-65980-R) (IP: J. Perera and L. Tolchinsky). [Table 2](#) provides a description of the aspects evaluated in each criterion.

Table 2

Description of analytical criteria

Criterion	Description
Content	Relevance and appropriateness of the text's content in relation to the proposed topic.
Organisation	Organisation of the content, with particular attention to how the different parts of the text are ordered and differentiated.
Cohesion	Connection between ideas within and between sentences is evaluated.
Vocabulary	Vocabulary appropriateness relative to the type of text is assessed, as well as lexical richness.
Correctness	Number of spelling, grammatical, and punctuation errors, and assessing their impact on comprehension.
Communicative impact	Clarity of the expressed point of view and the quality of the supporting evidence are assessed.

Teachers were asked to score each of these criteria, and to assign a global score to the text on a Likert scale from 1 to 5, where 5 was the highest score. Texts containing fewer than 40 words were assigned a score of 1 on all criteria because they were deemed too short for evaluation. Only two primary school participants were excluded from the analysis.

Assessment reliability was calculated on 20% of the texts using Cronbach's alpha coefficient. Significant agreement was reached among judges at all educational levels ($\alpha \geq .881$ for all scores) ([Field, 2009](#)).

Linguistic indicators of text quality

Morphosyntactic, lexical, and discourse-level measures, along with a measure of productivity, were obtained from the texts to analyse linguistic features previously linked to text quality.

- Productivity. Refers to the length of the text and was calculated based on the total number of words.
- Use of syntactic connectors. Refers to coordinating and subordinating conjunctions that connect clauses (intra-sentential connectivity devices). Their proportion was calculated relative to the total number of clauses.
- Use of discourse markers. Refers to connectivity devices that do not serve a syntactic function in clause predication but rather create supra-sentential links that contribute to textual cohesion (extra-sentential connectivity devices). Their identification followed [Martín-Zorraquino and Portolés-Lázaro \(1999\)](#), and their proportion was calculated relative to the total number of clauses.
- Lexical diversity. Refers to the variety of different words used in a text and was calculated using the D index, which —unlike other indices— is not affected by text length ([Malvern et al., 2004](#)).

- Lexical density. Refers to the proportion of content words in relation to the total number of words, providing insight into the information density of a text (Johansson, 2009).
- Lexical sophistication. Refers to the use of less common and more advanced words. It was calculated using the average length of words with semantic content, based on the premise that as word length increases, their occurrence decreases (Zipf, 1932).

Data analysis

The texts were transcribed in CHAT format and analysed using CLAN programmes from the CHILDES project (MacWhinney, 2000). The FREQ tool was used to determine the frequency of count variables, VOCB to calculate lexical diversity, and WDLN to obtain mean word length. The texts were segmented into clauses based on the criteria of Berman and Slobin (1994), as adapted for Spanish by Aparici (2010), and were morphosyntactically tagged using MOR and POST.

ANOVAs were conducted with educational level as between-subjects factor to analyse how external evaluation scores evolve and how the use of linguistic indicators of text quality varies across levels. Effect sizes were calculated using the eta-square value (η^2). Pairwise comparisons were conducted; Tukey's and Games-Howell's corrections were applied when the data did not meet the homoscedasticity criterion. Effect sizes were calculated using Cohen's d (Cohen, 1988).

To address the third objective, examining the relationship between linguistic indicators and text quality assessments, a series of Pearson correlations was conducted for each educational level. Regression analyses were then performed separately for each assessment criterion and for each educational level. Only indicators that significantly correlated with quality assessment scores were included as explanatory variables. The percentage of explained variance was determined using the adjusted R-squared coefficient (adjusted R^2), and the explanatory power of each variable was determined using the standardised beta coefficient (β).

RESULTS

The results have been divided into three subsections: the external evaluation of text quality by educational level; the use of linguistic indicators of text quality by educational level; and the relationship between external evaluations and linguistic indicators.

External evaluation of text quality by educational level

Educational level had a significant effect on all text quality scores, increasing with age. Table 3 shows the descriptive statistics, F values, and effect sizes. Results of pairwise comparisons are also included.

Table 3

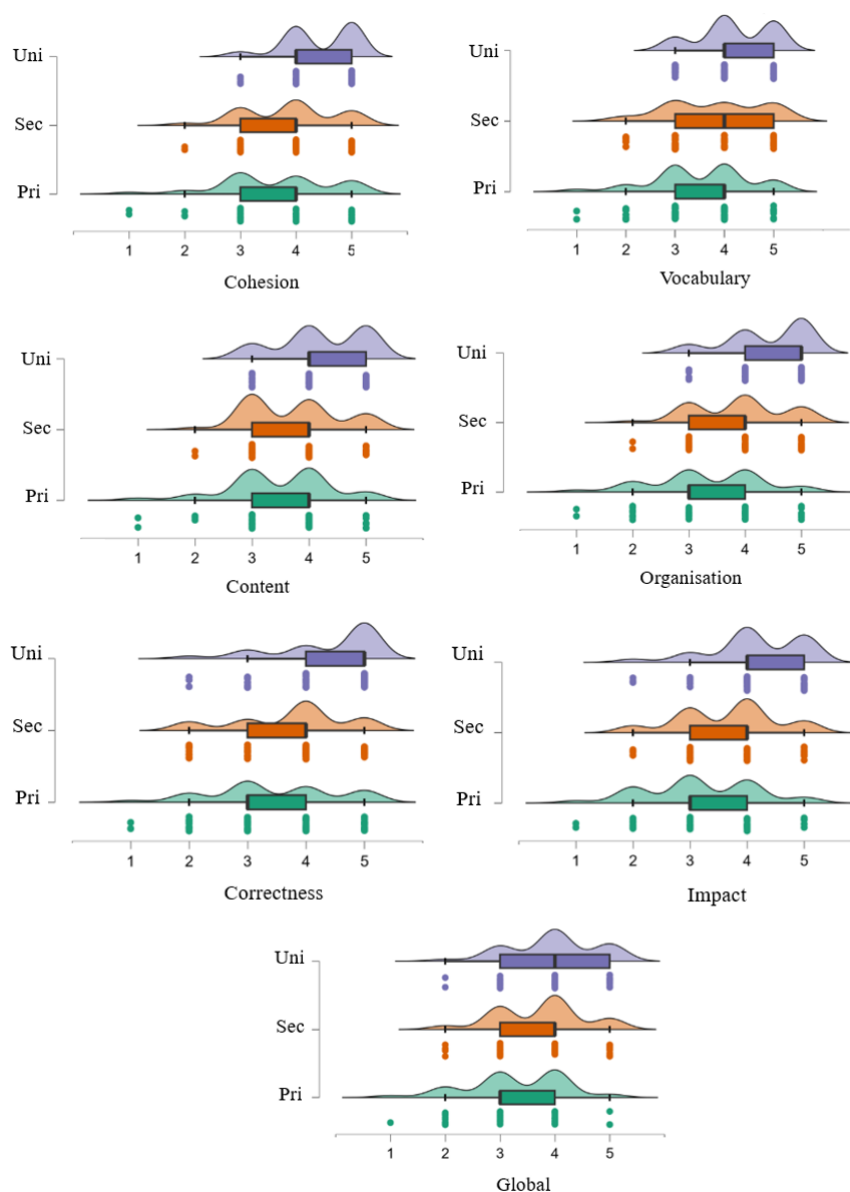
Descriptive statistics (mean (M) and standard deviation (SD)), F value, and effect size for text quality scores by educational level

Criteria	Educational level	M	SD	F	η^2
Content	Primary	3.48 ^a	0.90	14.448 ^{***}	0.121
	Secondary	3.72 ^b	0.80		
	University	4.22 ^{ab}	0.74		
Organisation	Primary	3.30 ^{ab}	0.97	28.486 ^{***}	0.214
Criteria	Educational level	M	SD	F	η^2
Cohesion	Secondary	3.88 ^{ac}	0.81	15.584 ^{***}	0.130
	University	4.39 ^{bc}	0.71		
	Primary	3.65 ^a	1.00		
Vocabulary	Secondary	3.87 ^b	0.83	9.901 ^{***}	0.087
	University	4.42 ^{ab}	0.63		
	Primary	3.52 ^a	0.97		
Correctness	Secondary	3.81 ^b	0.95	15.707 ^{***}	0.131
	University	4.20 ^{ab}	0.72		
	Primary	3.45 ^a	1.08		
Communicative impact	Secondary	3.74 ^b	0.95	20.181 ^{***}	0.162
	University	4.36 ^{ab}	0.89		
	Primary	3.19 ^{ab}	0.97		
Global score	Secondary	3.65 ^{ac}	0.85	12.351 ^{***}	0.106
	University	4.14 ^{bc}	0.81		
	Primary	3.27 ^{ab}	0.89		
	Secondary	3.73 ^a	0.78		
	University	3.97 ^b	0.78		
	Primary				

Values with the same superscripts are significantly different at $p < .05$;

$p < .001$.

Figure 1 presents text quality scores by educational level. Across all criteria, higher education participants scored higher on average than the other groups. Secondary school students also scored higher on average than primary school students. Variability was low at all educational levels, although it was slightly higher in primary school.

Figure 1*Distribution of text quality scores by educational level*

The results of the *post-hoc* tests indicated that primary school students received significantly lower scores than secondary school students only in organisation ($p < .001$; $d = 0.69$), impact ($p < .001$; $d = 0.54$) and the global score ($p < .001$; $d = 0.55$). They also scored significantly lower compared to higher education students across all evaluated aspects (content: $p < .001$; $d = 0.91$; organisation: $p < .001$; $d = 1.30$; cohesion: $p < .001$; $d = 0.93$; vocabulary: $p < .001$; $d = 0.76$; correctness: $p < .001$; $d = 0.94$; impact: $p < .001$; $d = 1.10$; global score: $p < .001$; $d = 0.85$). Secondary school students also received significantly lower scores than university students across all criteria (content: $p < .001$; $d = 0.61$; organisation: $p < .001$; $d = 1.30$; cohesion: $p < .001$; $d = 0.66$; vocabulary: $p < .05$; $d = 0.44$; correctness: $p < .001$; $d = 0.64$; impact: $p < .01$; $d = 0.56$), with the exception of the global score.

Use of linguistic indicators of text quality by educational level

Educational level had a significant effect on most linguistic indicators, with effect sizes ranging from small to medium. Table 4 shows the descriptive statistics, *F* values, and effect sizes. Results of pairwise comparisons are also included.

Table 4

Descriptive statistics (mean (M) and standard deviation (SD)), F value, and effect size for linguistic indicators

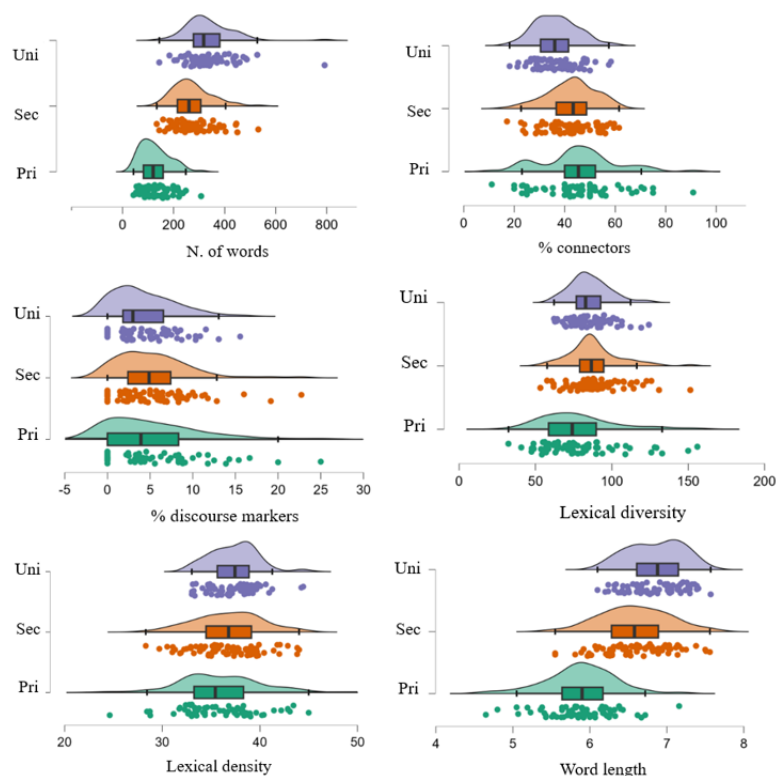
Indicators	Educational level	<i>M</i>	<i>SD</i>	<i>F</i>	η^2
No. of words	Primary ^{ab}	126.64	57.69	125.799 ^{***}	0.546
	Secondary ^{ac}	272.94	76.84		
	University ^{bc}	337.22	95.78		
Conectores	Primary ^a	45.47	15.26	10.596 ^{***}	0.092
	Secondary ^b	42.74	9.58		
	University ^{ab}	36.70	8.55		
Marcadores	Primary	5.23	5.48	1.312	0.012
	Secondary	5.31	4.36		
	University	4.20	3.62		
Diversidad léxica	Primary ^a	78.20	25.79	5.422 ^{**}	0.050
	Secondary ^a	88.63	17.20		
	University	85.62	12.83		
Densidad léxica	Primary	36.09	4.36	2.145	0.020
	Secondary	36.75	3.43		
	University	37.37	2.43		
Longitud palabra	Primary ^{ab}	5.88	0.45	95.612 ^{***}	0.478
	Secondary ^{ac}	6.58	0.44		
	University ^{bc}	6.85	0.36		

Values with the same superscripts are significantly different at $p < .05$;

$p < .01$;

$p < .001$.

Figure 2 presents the use of indicators by educational level. As shown, text length increases steadily with educational level. Moreover, for nearly all indicators, university students consistently outperform younger groups, except in the use of syntactic connectors, where the opposite is observed: primary school students produce a greater number of connectors than older students. Similarly, primary school students exhibit greater variability in performance than secondary and higher education students across all analysed measures.

Figure 2*Distribution of linguistic indicators by educational level*

Post-hoc tests showed a significant increase in text-length with educational level (primary vs. secondary: $p < 0.001$; $d = 1.86$; primary vs. university: $p < 0.001$; $d = 2.68$; secondary vs. university: $p < 0.001$; $d = 0.82$).

Regarding the use of connectors, both primary and secondary school students produced a significantly higher proportion of connectors than university students (primary vs. university: $p < .001$; $d = 0.77$; secondary vs. university: $p < .01$; $d = 0.53$), though no significant differences were observed between primary and secondary school groups. Unlike what was observed for syntactic connectors, no significant effect of educational level was found for discourse markers, indicating that their proportion does not change with age.

As for lexical measures, lexical diversity increased with age, although significant differences were observed only between primary and secondary school students ($p < .05$; $d = 0.55$). Meanwhile, average word length increased significantly with educational level (primary vs. secondary: $p < .001$; $d = 1.67$; primary vs. university: $p < .001$; $d = 2.32$; secondary vs. university: $p < .001$; $d = 0.64$). Finally, lexical density was not affected by educational level, remaining stable across groups.

Relationship between text quality assessment and linguistic indicators

The relationship between text quality assessments (analytical criteria and global evaluation) and linguistic indicators varies across educational levels. Except for correctness scores, which do not show significant correlations with any of the resources in any group, all other assessment criteria were associated with at least some linguistic indicator at one or more educational levels. [Figures 3, 4, and 5](#)

show the correlation results for each level. Based on these correlations, separate regression models were tested for each assessment within each group.

In primary school, all scores except correctness showed significant correlations with at least one linguistic resource. Text length significantly and positively correlated with most of the text quality criteria, as well as with global evaluation, indicating that longer texts tended to receive higher scores.

Figure 3

Correlation between scores and linguistic indicators in primary education

Global	0.583***	0.22	0.179	0.198	-0.15	0.083
Content	0.569***	0.19	0.148	0.335**	-0.184	0.128
Organisation	0.53***	0.171	0.153	0.111	-0.103	0.048
Cohesion	0.18	-0.05	0.277*	-0.054	-0.126	0.144
Vocabulary	0.372**	0.131	0.316*	0.118	-0.154	0.074
Correctness	-0.011	0.041	0.127	-0.009	-0.055	0.014
Impact	0.386**	0.178	0.171	0.081	-0.197	0.08
	N. of words	% connectors	% dis. markers	Lex. diversity	Lex. density	Word length

Note. * $p < .05$; ** $p < .01$; *** $p < .001$

In fact, the only indicator that correlated significantly with the global score was the number of words ($r = .583$). Therefore, the regression model tested included productivity as the only predictive variable. This model proved to be significant, $F_{(1,63)} = 32.466$; $R^2 = .331$; $p < .001$, explaining 33.10% of the overall score. Text length contributed significantly and positively to the model ($\beta = 0.583$; $p < .001$).

Regarding the content criterion, it correlated significantly with text length ($r = .569$) and lexical diversity ($r = .335$). Thus, the regression model included both indicators as predictive variables. The model was also significant, $F_{(2,61)} = 17.929$; $R^2 = .350$; $p < .001$, accounting for 35% of the variance. Both lexical diversity ($\beta = 0.267$) and text length ($\beta = 0.512$) contributed positively to the model ($p < .05$).

Regarding organisation, only text length correlated significantly ($r = .530$). The regression model was significant, $F_{(1,63)} = 24.674$; $R^2 = .281$; $p < .001$, explaining 28.1% of the variance. Text length contributed significantly and positively to the model ($\beta = 0.530$; $p < .001$).

With regard to cohesion scores, these only correlated significantly with the proportion of discourse markers ($r=.277$). The model was significant, $F_{(1,63)}=5.237$; $R^2=.062$; $p<.025$, explaining 6.2% of the variance in cohesion scores. The proportion of markers contributed significantly to the model ($\beta=0.277$; $p=.025$).

With respect to vocabulary, two significant correlations were found with text length ($r=.372$) and with the proportion of discourse markers ($r=.316$). The regression model was also significant, $F_{(2,63)}=5.280$; $R^2=.118$; $p=.008$, and explained 11.8% of the variance in scores, although only text length contributed significantly ($\beta=0.361$; $p=.003$).

Finally, the scores given for communicative impact only correlated with text length ($r=.386$). The tested model was significant, $F_{(1,63)}=11.008$; $R^2=.149$; $p=.002$, explaining 14.9% of the variance. Text length contributed significantly ($\beta=0.386$; $p=.002$).

In secondary school, fewer significant correlations were observed between linguistic resources and text quality scores. Significant correlations emerged only for the global score, and for the content and vocabulary scores.

Figure 4

Correlation between scores and linguistic indicators in secondary education

Global	0.251*	0.084	0.041	0.179	0.125	0.162
Content	0.03	-0.137	0.16	0.253*	0.297**	0.248*
Organisation	0.061	0.041	0.071	-0.01	-0.027	0.035
Cohesion	0.068	0.058	0.061	0.011	-0.155	-0.009
Vocabulary	0.281*	0.071	-0.097	0.187	0.099	0.117
Correctness	0.08	-0.017	0.021	0.184	0.026	0.065
Impact	0.038	0.01	0.101	0.184	0.044	0.073
	N. of words	% connectors	% dis. markers	Lex. diversity	Lex. density	Word length

Note. * $p<.05$, ** $p<.01$, *** $p<.001$

In relation to the global score, as in primary education, text length was the only indicator that correlated significantly with this variable ($r=.251$). The tested model was also significant, $F_{(1,76)}=5.123$; $R^2=.051$; $p=.026$, but it only explained 5.10% of the variance. Text length significantly contributed to the model ($\beta=0.251$; $p=.026$).

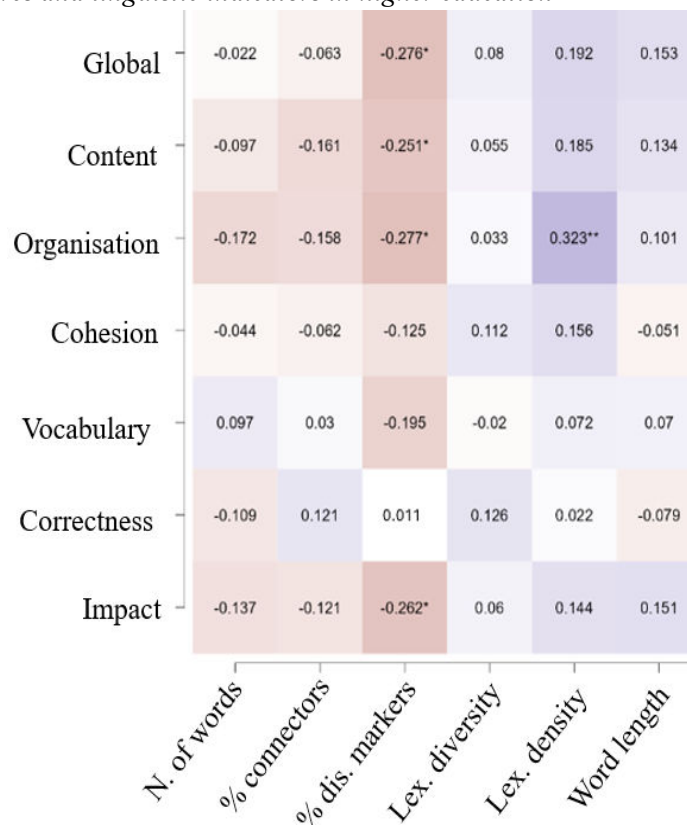
Content scores correlated significantly with all lexical indicators: lexical diversity ($r=.297$), lexical density ($r=.297$), and average word length ($r=.248$). The model was significant, $F_{(3,74)}=3.420$; $R^2=.086$; $p<.001$, explaining 8.6% of the variance; however, none of these indicators contributed significantly to the model.

Finally, vocabulary scores correlated significantly with text length ($r=.281$). The model tested was significant, $F_{(1,76)}=6.512$; $R^2=.067$; $p=.013$, accounting for 6.7% of the variance. The number of words contributed significantly to the model ($\beta=0.281$; $p=.013$).

As in secondary education, fewer significant correlations were found in higher education between linguistic resources and text quality scores. Significant correlations were observed for the global score, and for the content, organisation, and impact scores. In contrast with the younger age groups, text length did not correlate significantly with any of the scores in higher education.

Figure 5

Correlation between scores and linguistic indicators in higher education



Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

With respect to the global score, the only significant correlation was with the proportion of discourse markers, which was negative ($r = -.298$). The model was significant, $F_{(1,67)} = 6.508$; $R^2 = .089$; $p = .013$, explaining 8.90% in global scores. The proportion of discourse markers contributed negatively to the model ($\beta = -0.298$; $p = .013$).

Content scores also correlated negatively with the proportion of discourse markers ($r = .271$). The tested model was significant, $F_{(1,67)} = 5.329$; $R^2 = .074$; $p < .001$, accounting for 7.4% of the variability in content scores. Again, the proportion of discourse markers also contributed negatively to the model ($\beta = -0.271$, $p = .024$).

Regarding organisation scores, a negative correlation was observed with the proportion of discourse markers ($r = -.293$), and a positive correlation with lexical density ($r = .332$). The model was also significant, $F_{(2,66)} = 6.122$; $R^2 = .156$; $p < .004$, explaining 15.6% of the variance in organisation scores. Both lexical density ($\beta = 0.275$) and the proportion of discourse markers ($\beta = -0.222$) contributed significantly ($p < .05$), albeit in opposite directions: lexical density had a positive effect, while the proportion of discourse markers had a negative effect.

Finally, for communicative impact, a negative correlation with the proportion of discourse markers was again observed ($r=-.273$). The model was significant, $F_{(2,63)}=7.333$; $R^2=.118$; $p<.008$, explaining 11.8% of the variance. Once again, the proportion of discourse markers contributed negatively to the model ($\beta=-0.273$; $p=.023$).

DISCUSSION

This study analysed the quality of analytical texts written by primary, secondary, and higher education students. First, it examined the text quality assessments made by expert teachers at these educational levels; second, it explored the development of a set of linguistic indicators of text quality across educational levels. In addition, the relationship between these two types of assessments was analysed to identify the indicators that best explain teachers' evaluations.

Regarding the first objective, which focused on variations in teachers' assessments, the results indicate that the highest-rated texts were those produced by university students, except for the global score, where no significant differences were observed between secondary and higher education. This finding could indicate that, although university students outperform younger students in specific linguistic and textual dimensions, teachers prioritise more general aspects, such as clarity or coherence, when assigning a global evaluation. Alternatively, this pattern may reflect teachers' adjustments of expectations and, thus of their evaluations to the educational level being assessed (Cuberos-Vicente, 2019; Salas et al., 2016). Primary and secondary school students received similar ratings for most dimensions evaluated, with the exception of impact, organisation, and the global score, where primary school students received lower evaluations. This suggests that certain macrostructural and rhetorical aspects typical of this type of text begin to be explicitly assessed from secondary school onwards. At this stage, students' texts increasingly exhibit features typical of expository genres, whose production relies on linguistic resources used that differ from those used in earlier-acquired genres, such as narrative texts (Aparici, 2010; Berman & Verhoeven, 2002).

With respect to our second objective, to analyse the development of a set of linguistic indicators of text quality, the results show differences in nearly all the linguistic resources analysed across educational levels, although each indicator follows a distinct developmental pattern. Consistent with previous research, students at higher educational levels produce longer texts and use a more sophisticated and diverse vocabulary. These findings provide further evidence that text length, lexical diversity, and lexical sophistication—often operationalised as the use of longer words—are indicators of academic development (Aparici et al., 2021; Berman and Nir-Sagiv, 2010; Crossley and McNamara, 2010). Despite the increase in productivity and the use of more sophisticated and diverse vocabulary associated with educational advancement, lexical density remains stable. This pattern has been reported in previous studies in Spanish (Aparici et al., 2021; Cuberos-Vicente, 2019; Salas et al., 2016) and Catalan (Llauradó and Tolchinsky, 2013), but it contrasts with results from English, Hebrew, and Swedish (Berman and Nir-Sagiv, 2010; Strömquist et al., 2002). These divergent trends could reflect typological differences between languages (Johansson, 2009), such as the degree to which rely on independent function words rather than on bound morphemes to encode grammatical relations. In the case of connective devices, distinct patterns of development were identified. While the proportion of discourse markers remains consistent across educational levels, the proportion of syntactic connectors decreases with age. These findings extend the developmental trend documented in a parallel study for Catalan (Rosado et al., 2021) and may be attributable to several factors. Firstly, the decrease in the use of syntactic connectors among university students is consistent with Crossley et al.'s (2011) claim that cohesion is not necessarily achieved through the abundant use of connectors. Therefore, this downward trend—or the absence of developmental differences—could indicate university students rely on alternative cohesive mechanisms. In fact, the optional nature of discourse markers, which is one of the rhetorical options available to writers, could partly account for this behaviour. Another possible explanation is that the present study did not distinguish between the different discursive functions (e.g., structuring or modalisation) performed by these markers, which could reveal age-related

differences (Rosado et al., 2021). In addition, non-conventional discourse markers —i.e. multi-word units that perform the same functions as traditional discourse markers but are not categorised as such (Cuenca, 2013; Galiana-Bea et al., 2024)— were not analysed. Nor were non-canonical uses of the analysed markers identified, that is, grammatically or semantically inappropriate uses within the discourse context.

Regarding the third objective —identifying which linguistic resources explain variations in teachers' assessments of text quality— the results show clear differences across educational level. In primary school, text length emerges as the indicator of text quality par excellence, accounting for variations in the scores for all assessed aspects, except correctness. Consistent with previous studies (Cuberos, 2019; McMaster and Espin, 2007), this finding suggests that producing longer texts entails the use of additional linguistic resources that enable writers to articulate more precise arguments and reasoning. At this level, a positive association was also observed between lexical diversity and content scores, suggesting that writers who use a more varied vocabulary tend to engage more deeply with the topic. Furthermore, increases in the use of discourse markers were associated with higher vocabulary and cohesion scores, which corroborates previous findings that more cohesive texts tend to receive higher evaluations (Aparici et al., 2021). The use of discourse markers, though, is not associated with organisation scores, which may suggest that teachers value their cohesive function but may not view them as central to the overall organisation of the text. Surprisingly, the use of syntactic connectors, traditionally linked to textual cohesion (Hyland, 2002), was not associated with either cohesion or organisation scores. These results highlight the need for future research to incorporate functional criteria that allow for a more nuanced discussion of the discursive value of these markers (Rosado et al., 2021).

In secondary school, a similar pattern emerges regarding the relationship between text length and teacher assessments: content, vocabulary, and global scores all increase as text length increases. These results suggest that productivity remains a key factor in assessing text quality at this stage of schooling (Berman and Nir-Sagiv, 2010; Salas et al., 2016). At this level, all lexical measures are positively associated with content scores, and lexical diversity also contributes to global evaluations. This indicates that students' lexical repertoire is particularly relevant to the assessment of text quality during secondary education. However, neither the use of discourse markers nor the use of syntactic connectors is associated with teacher scores, reinforcing the need for further research on the characterisation of connective devices based on functional criteria.

In higher education, however, productivity is not directly associated with teacher evaluations. This suggests that, once the minimum length needed to articulate a viewpoint is achieved, evaluators can begin to focus on aspects other than text length when judging writing quality. Surprisingly, a greater use of discourse markers is negatively associated with most text quality scores, which leads us to hypothesise that some of these markers are semantically or discursively inappropriate. Such non-canonical uses, noted in earlier studies (López-Ferrero and Atienza-Cerezo, 2006; Rosado et al., 2021), warrant detailed examination in future research. At this educational level, a positive association was also found between lexical density and text organisation, suggesting that structural organisation may be facilitated by an increased information density (Johansson, 2009).

Overall, teachers do not seem to apply the same evaluation criteria across educational levels, although certain aspects—such as productivity and lexical diversity—remain stable predictors of text quality in primary and secondary education.

One aim for future research is to provide a more comprehensive analysis of the nature of the discourse markers used in the texts. This would enable us to more fully understand the discrepancies between the forms used and their contextual functions, in line with the work of Alonso-Chacón (2019) and Tolchinsky et al. (2021). Furthermore, we consider it advisable for teachers to evaluate texts from different educational levels, rather than limiting themselves to those from the educational level at which they teach. This approach would help clarify how perceptions of text quality may vary depending on teachers' experience and educational context. It could also yield valuable insights for designing and implementing differentiated teaching strategies for written language instruction.

CONCLUSIONS

his study confirms that the quality of written academic texts develops progressively across educational levels, both in terms of the use of linguistic resources and of teacher evaluation. In primary and secondary education, productivity and lexical diversity emerge as the most relevant indicators, whereas in higher education, the inappropriate use of discourse markers can negatively influence perceptions of text quality. These findings underscore the need to develop assessment tools that are aligned with each educational level, as well as the need for a more in-depth functional analysis of the linguistic resources used by the students, particularly at more advanced levels.

CONTRIBUTION OF THE AUTHORS

Rocío Cuberos-Vicente: Formal analysis; Conceptualisation; Data curation; Writing - original draft; Writing - revision and editing; Research; Methodology; Resources; Supervision; Validation; Visualisation; Fund acquisition.

Elisa Rosado: Project management; Conceptualisation; Writing - original draft; Writing - revision and editing; Research; Methodology; Resources; Supervision; Validation.

Verónica Martínez: Formal analysis; Conceptualisation; Data curation; Writing - original draft; Writing - review and editing; Research; Methodology; Supervision; Validation; Visualisation.

Melina Aparici: Project management; Conceptualisation; Writing - review and editing; Research; Methodology; Resources; Supervision; Validation; Fundraising.

FUNDING

his study was funded by the Ministry of Science, Innovation and Universities (PID2020-119555GA-I00; I.P.: M. Aparici) and by a Margarita Salas postdoctoral grant awarded to Rocío Cuberos-Vicente by the University of Barcelona and funded by Next Generation EU and the Ministry of Universities.

DECLARATION

A previous version of this study was presented at the 33rd International Congress of the Spanish Association of Speech Therapy, Phoniatrics and Audiology and the Ibero-American Association of Speech Therapy held in Santander from 28 to 30 September 2023. The abstract was published in the special volume 43(1) of the *Journal of Speech Therapy, Phoniatrics and Audiology*. <https://doi.org/10.1016/j.rlfa.2023.100372>.

REFERENCES

- Alonso-Chacón, P. J. (2019). *Uso de marcadores discursivos en el discurso académico oral y escrito de estudiantes universitarios costarricenses* [Tesis doctoral, Universitat de Barcelona]. <https://hdl.handle.net/2445/151342>
- Aparici, M. (2010). *El desarrollo de la conectividad discursiva en diferentes géneros y modalidades de producción* [Tesis doctoral, Universitat de Barcelona].
- Aparici, M., Cuberos, R., Salas, N., & Rosado, E. (2021). Linguistic indicators of text quality in analytical texts: developmental changes and sensitivity to pedagogical work. *Journal for the Study of Education and Development*, 44(1), 9-46. <https://doi.org/10.1080/02103702.2020.1848093>
- Aparici, M., Rosado, E., Vilar, H., Cuberos, R., & Tolchinsky, L. (2024). The influence of students' linguistic condition, school level, and pedagogical input on analytical essay features. *Frontiers in Language Sciences*, 3, 1480422. <https://doi.org/10.3389/flang.2024.1480422>

- Berman, R. A., & Nir-Sagiv, B. (2010). The lexicon in writing–speech–differentiation. Developmental perspectives. *Written Language and Literacy*, 13(2), 183-205. <https://doi.org/10.1075/wll.13.2.01ber>
- Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Erlbaum. <https://doi.org/10.4324/9780203773512>
- Berman, R. A., & Verhoeven, L. (2002). Cross-linguistic perspectives on the development of text production abilities: Speech and writing. *Written Language and Literacy*, 5(2), 1-43. <https://doi.org/10.1075/wll.5.1.02ber>
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Erlbaum. <https://doi.org/10.4324/9780203771587>
- Crossley, S. A., & McNamara, D. S. (2010). Cohesion, coherence, and expert evaluations of writing proficiency. En S. Ohlsson y R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 984-989). Cognitive Science Society. <https://escholarship.org/uc/item/6n5908qx>
- Crossley, S. A., Weston, J., 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. <https://doi.org/10.1177/0741088311410188>
- Cuberos-Vicente, R. (2019). *Indicadores léxicos de calidad textual en español nativo y no nativo* [Tesis doctoral, Universitat de Barcelona]. <https://diposit.ub.edu/dspace/handle/2445/178686>
- Cuenca, M. J. (2013). The fuzzy boundaries between discourse marking and modal marking. En L. Degand, B. Cornillie y P. Pietrandrea (Eds.), *Discourse markers and modal particles. Categorization and description* (pp. 181-216). John Benjamins. <https://doi.org/10.1075/pbns.234>
- Field, A. (2009). *Discovering statistics using SPSS (and sex and drugs and rock'n'roll)* (3ª Ed.). SAGE Publications.
- Galiana-Bea, P., Gras-Manzano, P., Rosado, E., & Mañas-Navarrete, I. (2024). Marcadores discursivos y otros mecanismos de marcación discursiva. Una propuesta holística para el análisis de narraciones orales en español como lengua extranjera. *Rilce*, 40(3), 937-69. <https://doi.org/10.15581/008.40.3.937-69>
- Gavari-Starkie, E. I., & Tenca-Sidotti, P. (2017). La evolución histórica de los Centros de Escritura Académica. *Revista de Educación*, 378, 9-29. <https://doi.org/10.4438/1988-592X-RE-2017-378-359>; <https://www.educacionfpydeportes.gob.es/revista-de-educacion/gl/dam/jcr:6b2bc342-a3f0-48dd-8148-0b3ee1a8ff5/01gavari-pdf.pdf>
- Hyland, K. (2002). *Teaching and researching writing*. Longman/Pearson. <https://doi.org/10.4324/9781003198451>
- Johansson, V. (2009). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Working Papers Lund University*, 53, 61-79. <https://journals.lub.lu.se/LWPL/article/view/2273>
- Llauradó, A., & Tolchinsky, L. (2013). Growth of text-embedded lexicon in Catalan: From childhood to adolescence. *First Language*, 33(6), 628-653. <https://doi.org/10.1177/0142723713508861>
- López-Ferrero, C., & Atienza-Cerezo, E. (2006). Las conjunciones paratácticas en el Corpus 92. En E. Bernal y J. A. DeCesaris (Eds.), *Palabra por palabra: estudios ofrecidos a Paz Battaner* (pp. 147-160). Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada. <http://hdl.handle.net/10230/23683>
- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3ª Ed.). Erlbaum.

- Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development. Quantification and assessment*. Palgrave MacMillan. <https://doi.org/10.1057/9780230511804>
- Martín-Zorraquino, M. A., & Portolés-Lázaro, J. (1999). Los marcadores del discurso. En I. Bosque, & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (Vol. 3, pp. 4051-4213). Espasa Calpe.
- McMaster, K., & Espin, C. (2007). Technical features of curriculum-based measurement in writing. *The Journal of Special Education*, 41(2), 68-84. <https://doi.org/10.1177/00224669070410020301>
- McNamara, D. S., Crossley, S. A., & McCarthy, P. (2010). Linguistic features of writing quality. *Written Communication*, 27(1), 57-86. <https://doi.org/10.1177/0741088309351547>
- Rosado, E., Mañas-Navarrete, I., Yúfera-Gómez, I., & Aparici-Aznar, M. (2021). El desarrollo de la escritura analítica: aprender a enlazar la información, aprender a posicionarse. *Pensamiento Educativo*, 58, 1-18. <https://doi.org/10.7764/PEL.58.2.2021.10>
- Salas, N., Llauradó, A., Castillo, C., Taulé, M., & Martí, M. A. (2016). Linguistic correlates of text quality from childhood to adulthood. En 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. https://doi.org/10.1007/978-3-319-21136-7_18
- Schleppegrell, M. J. (2004). *The language of schooling. a functional linguistics perspective*. Lawrence Erlbaum. <https://doi.org/10.4324/9781410610317>
- Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. En D. R. Olson y N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp. 112-133). Cambridge University Press. <https://doi.org/10.1017/CBO9780511609664.008>; <http://nrs.harvard.edu/urn-3:HUL.InstRepos:11654980>
- Strömquist, S., Johansson, V., Kriz, S., Ragnarsdóttir, H., Aisenman, R., & Ravid, D. (2002). Toward a cross-linguistic comparison of lexical quanta in speech and writing. *Written Language and Literacy*, 5(1), 45-67. <https://doi.org/10.1075/wll.5.1.03str>
- Tolchinsky, L., Aparici-Aznar, M., & Rosado, E. (2017). Escribir para pensar y persuadir. *Textos de Didáctica de la Lengua y la Literatura*, 76, 14-21. <https://hdl.handle.net/2445/122119>
- Tolchinsky, L., Aparici, M., & Vilar, H. (2021). Macro- and micro-developmental changes in analytical writing of bilinguals from elementary to higher education. *International Journal of Bilingual Education and Bilingualism*, 25(7), 2511-2526. <https://doi.org/10.1080/13670050.2021.1923643>
- Tolchinsky, L., Rosado, E., & Aparici, M. (2023). Internal and external appraisals of analytical writing. A proposal for assessing development and potential improvement. *International Review of Applied Linguistics in Language Teaching*, 62, 5-36. <https://doi.org/10.1515/iral-2023-0012>
- Toulmin, S. E. (2002). *The uses of argument*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511840005>
- Uccelli, P. (2023). Midadolescents' language learning at school: Toward more just and scientifically rigorous practices in research and education. *Language Learning*, 73(2), 182-221. <https://doi.org/10.1111/lang.12558>
- Uccelli, P., Deng, Z., Phillips-Galloway, E. P., & Qin, W. (2019). The role of language skills in midadolescents' science summaries. *Journal of Literacy Research*, 51(3), 357-380. <https://doi.org/10.1177/1086296X19860206>
- Uccelli, P., Dobbs, C. L., & Scott, J. (2012). Mastering academic language: organization and stance in the persuasive writing of high school students. *Written Communication*, 30(1), 36-62. <https://doi.org/10.1177/0741088312469013>
- Zipf, G. K. (1932). *Selected studies of the principle of relative frequency in language*. Harvard University Press. https://pure.mpg.de/rest/items/item_2407800_3/component/file_2459540/content