Improving Academic Writing with a Method for Text Revision Supported by Text Mining

https://doi.org/10.3991/ijet.v17i21.31249

Simone de Oliveira[1], Eliseo Reategui[1](✉), Patrícia S. C. C. Barcellos[1], Marcio Bigolin[1], Michel Carniato[2]
[1]PGIE, Federal University of Rio Grande do Sul, Porto Alegre, Brazil
[2]PUCRS, Porto Alegre, RS, Brazil
eliseoreategui@gmail.com

Abstract—In spite of the undisputed importance of writing, especially in the learning context of higher education, we still find many students with difficulties to write, present arguments, and keep their writing coherent. Therefore, this research aims to investigate how a learning tool with text mining functions can support students in the process of academic writing, based on certain criteria for text analysis. A quasi-experimental study with control and experimental groups was carried out in order to evaluate the tool's capacity to help students revise and improve their texts. The experimental group had access to the mining tool, which guided them in the revision of global features of their texts according to four principles: continuity, congruency, progression and non-contradiction. A group of lecturers corrected the students' tasks using the same correction grid. Results showed that the students who used the mining tool achieved higher grades when compared to the students who did not use it. Providing one-to-one support for the revision of certain text facets at the global level has been one of the main contributions of this research.

Keywords—academic writing, text mining, revision

1 Introduction

Computers have changed the way people write from a very early age [1]. Technology influences cognitive processes involved in writing and often enables people to write more and better when using computers [2][3]. In higher education, writing and the mastery of language play a major role as students need to understand academic expectations and continually use written language to communicate what they've learned [4]. Therefore, understanding how students can use methods and tools to improve their text production skills is an important research goal. Graham and Harris [5] stressed the need for the development of a supportive and pleasant writing environment for the students, in which writing activities could be performed, especially with the support of word processors. In this context, technology has been shown to be able to help students improve their writing skills [6]. However, with the increasing workload teachers have to deal with, novel demands emerge. For example, the need for tools that can support writing

instruction and help students in their actual text production and revision [7]. On this basis, Strobl et al. [8] presented a review of tools and pedagogies available to support writing in secondary and higher education. The authors identified that technological support at the micro-level is more frequent, with tools and methods to help students revise texts mostly in terms of grammar and spelling. The same has been observed with the use of ICT (Information and Communication Technologies) in writing instruction in K-12, where digital tools are employed mostly for editing purposes, and automated support is provided mostly at the micro-level [9]. Tools to support the development of writing strategies at the macro-level are less frequent. These tools tackle the problem of helping students self-monitor their writing abilities so that they can improve the formal and the argumentative structure of their texts. The development of such a tool has been the main goal of the work presented in this article. It follows a specific research approach to academic writing that focuses on self-regulation to support students in revision tasks. The development of self-regulation skills, in particular, operate in the adaptation of strategic choices to specific requirements of given academic writing tasks [10]. Furthermore, it has been indicated that learning to write academically requires knowledge about academic discourse, which stresses the need to be able to recognize and reproduce academic genres [11].

Previous research has also shown the positive effects of multimodal components on the writing process [12][13]. Visual representations such as concept maps have also been shown to benefit students in writing tasks in Science [14]. Providing students with tools to illustrate their organization of knowledge and to express visually their ideas is an approach that has been supported by different authors [15][16]. These research works have focused mainly on the use of graphic applications that allow users to create their own representation of propositions about facts. The approach presented in this article has also focused on providing users with functions to represent ideas in a graphical way. However, our specific research problem was whether the tool could help students improve their writing in a global level by questioning them about structural aspects of their work. The assumption was that students could improve their writing by reviewing macro aspects of their own work with the support of graphical representations automatically extracted from their texts, using a known framework for coherence analysis defined in [17].

The tool designed in this research makes use of text mining functions to extract and represent relevant terms and relationships from students' texts, it then uses graphical representations of the extracted information to pose questions that make learners reflect and review their writing. In the study presented in this article, this guided revision has shown to get students to make text changes that resulted in improved manuscripts.

2 A theoretical approach to writing with the support of technology

Studies from the 1960s proposed a decomposition of the writing process in three stages: pre-writing, writing, post-writing [18]. These stages were represented as if they were sequential and separated in time. Linear models had a unidirectional shape, which did not effectively represent the way in which authors write. Even so, such models

served as a basis for the development of more complex ones, some of them highly influential on the way one understands the cognitive process of writing to date. These models also represent the writing process of experienced and inexperienced writers, which is an interesting aspect when one wants to help novice writers become more proficient [19].

From a selection of writing process models, Strobl et al. [8] identified that writing subtasks are frequently divided in prewriting, planning, drafting, reviewing, and editing. One of the most influential cognitive models of writing has been developed by Flower and Hayes [20] from verbal protocols. The model represented the writing process in 3 main categories: planning, translation and revision. The planning module was related to the process of understanding the context and organizing ideias to be expressed in the text. But according to the authors, planning is not just a detailed plan, but an internal representation of the knowledge that the writer has and uses in written production. The translation activity is the transposition of ideas into actual language, the translation of a meaning into linear writing, under the guidance of the writing plan. The revision activity is divided into two subcategories: reading, which concerns the identification of problems in the text; and editing, which establishes the real changes to be made in the text. Chenoweth & Hayes [21] call this process text production, in their model of revision.

In this sense it is interesting to highlight Becker's [22] conclusion that, in the past, revision was seen as a fairly simple reviewing task, which occurred at the end of the writing process. However, research in the field has demonstrated that revision is a highly complex operation, an essential activity that builds skill levels and promotes writing expertise. Revising is not a sub strategy of writing in the same way as planning, transcribing, and reviewing is. It is a rewriting process that takes place when the already-existing text is read for evaluation. Hayes, Flower, Schriver, Stratman and Carey [23] state that writing has "privileged access" to the intentions that gave rise to the text. This can be both a help and an obstacle. It helps when only writers know if a particular word or phrase exactly matches their intention. But it can be an obstacle when the writers' knowledge of their own intentions can blind them to certain flaws in their texts. From the research results reported in [23] it is evident that diagnostic skills are often the most important factors in successfully revising texts, both on local and global levels. While at the local level writers look for spelling and grammar mistakes, at the global level, they look for coherence problems. Jones [24] explains the concept by stating that, in a coherent text, ideas flow smoothly from one sentence to the next, from one paragraph to another, in a logical order. An idea that is unrelated to the general topic or the local context challenges coherence. A known framework for evaluating coherence has been defined by Charolles [17], in which a text should meet four rules to achieve coherence: continuity, congruency, progression and non-contradiction.

- Continuity: it is manifested through the repetition of ideas and concepts, giving consistency to the document as a unified whole.
- Congruency: clear connections between the main idea of the text and secondary ones are present, as well as relationships among the secondary ideas.
- Progression: each part of the text is built on a prior one, producing a logical structure. New information is added to concepts that are repeated along the text in a way that it progresses without restricting itself to the same ideas from beginning to end.

- Non-contradiction: the ideas presented do not contradict each other, i.e. the writer cannot contradict a statement he made in the same text.

When it comes to supporting proofreading and reviewing at the local level, there are several established text editors and other online tools that leave markings in sentences or words to highlight grammar and misspelling problems. However, the number of tools is more limited to help learners write more coherent texts, which requires text analyses at the global level.

Li and Tsai [25] propose a text structure learning system in which the students, after writing, assemble a mental map of the text and the teacher can edit this mental map by asking questions that need to be referenced in the text. Hyerle [26] showed how different types of visual tools could help students and teachers represent information and communicate with others. Providing tools to allow users to structure ideas in the context of writing is an approach that has been supported in previous research, most of it focused on pre-writing [27] and, more specially, second-language learning [28][29]. In this article, the idea has also been to provide support to students at the global level of writing. However, the efforts here were not focused on the pre-writing phases of the process, but in the reviewing phase through the use of a text mining tool presented in the next section.

3 Text mining in education

A recent literature review [30] showed that the increasing amount of unstructured textual material involved in learning processes have drawn researchers to investigate the use of text mining in the analysis of different learning sources, such as: online assignments, essays, chats, emails and posts in discussion forums, social networks and blogs.

Researchers have also evaluated how text mining used for the analysis of student responses from surveys could yield relevant information for management purposes [31], contrasting the results obtained through the mining of students' opinions about teacher leadership with those of human raters [32]. Text mining has also been used to classify resources provided to students in online learning courses [33] and to analyze blended learning strategies to identify better practices in terms of model selection and infrastructure preparation [34]. It has also been employed in the analysis of students' writings, from shallow features and error detection [35] to the identification of mistakes in grammar, linguistic usage, and style [36]. Furthermore, text mining has shown to be useful in formative assessment and letting learners and instructors visualize results, providing an alternative solution to evaluate learners' performance throughout the learning process [37].

As for the use of text mining to support writing, graphical representations of texts have been used to help students in summarization tasks [38]. Another system targeted to support students in writing tasks is OpenEssayist [39], a real-time learning analytics tool that provides automated feedback to students, getting them to reflect on their work as a means to improve their writing through understanding of the requirements of academic text production.

This article presents a study in which the text mining system called Sobek[1] [40] was integrated into *TexRev*, a tool designed to support writing, especially the revision of academic texts. The integrated text mining features provided students with graphical representations of the most relevant terms and relationships found in their texts. The mining algorithm used operates in three stages: the first one consists in identifying the relevant concepts in the text; the second step is the identification of relationships among these concepts; the last step concerns the representation of the information extracted in the form of graphs. These graphs can be thought of as graphic organizers, which have been widely used to help students and teachers represent information and communicate with others [41][26].

To illustrate the tool's graphical representation of information extracted from texts, Figure 1 shows a graph obtained from a Wikipedia text about Global warming[2].

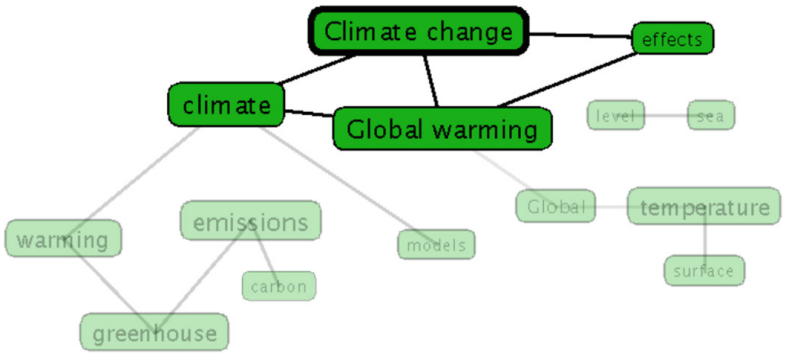

Fig. 1. Graph extracted from a Wikipedia text on Global Warming

The larger the node in the graph, the higher its relative frequency in the text when compared to those of other nodes. The graph produced has a set of functionalities that allows the user to find further information about each node and relationship. For instance, when clicking on a given node, its frequency is displayed below the graph, as well as fragments of the text in which the term appears.

4 Methods

Our focus in this research has been to provide students with a technological tool that would be able to help them in writing activities, addressing mainly the revision of texts from a macro-structure perspective. We started by integrating Sobek [40] into *TexRev*, with the goal of providing students with a guided strategy for text revision based on information extracted from their own texts. The idea here was to present students with graphical representations of their complete texts, as well as of separate parts of them. These visual representations were combined with specific questions used by TexRev to guide students in their text analysis. It is important to highlight that the tool was

[1] http://sobek.ufrgs.br/#/

[2] https://en.wikipedia.org/wiki/Global_warming

specifically built to support learners in writing expository essays, a genre that aims at presenting opinion-free information on a given topic, structured here in three sections: Introduction, Development and Conclusion. Therefore, the tool was not targeted to help experienced writers, or to support writing of other genres such as poetry, romance, or other fiction genres.

A quasi-experimental study with control and experimental groups was carried out in order to evaluate the tool's capacity to help students revise and improve their work. The study was conducted in a private higher education institution in the south of Brazil, in a region with a high human development index. Four classes of the same introductory university course took part in the study, totalling 728 students. The large majority of students were in the age range of 18–23 years old. The course introduced concepts related to personal and professional development, placing writing as a fundamental skill. However, only 496 students finished all the tasks needed for their complete participation in the experiment. Therefore, this was the number of participants considered in the experiment, with the following final distribution of students in the control and experimental group: control group (n = 211), experimental group (n = 285). To verify the homogeneity of the groups, the performance of the students in seven writing activities prior to the experiment was analyzed. The average performance of both groups had no significant difference, with the experimental group attaining an average grade of 8,49 (standard deviation = 1,95) and the control group attaining 8,48 (standard deviation = 1,86).

Next, the students of the experimental and the control group had to write an expository essay about the topic "Entrepreneurship", structured in three sections: Introduction, Development, and Conclusion. The students had two weeks to complete the task. Both groups received the same material with writing instructions about the construction of an essay, but the experimental group had access to TexRev, the tool which incorporated text mining functions with embedded questions to help learners in the reviewing task. Considering Flower & Hayes' [20] influential cognitive models of writing and its tasks' subset (i.e. planning, translating and reviewing), this work has focused mainly on the use of TexRev to help students in the reviewing phase. The goal here has been to provide support for text revision at the global level, following the framework for text coherence evaluation detailed in section 2 [17], focused on the factors: ***continuity, congruency, progression and non-contradiction*.**

The idea has been to ask students to analyze graphs that were extracted from their texts concerning different parts of their manuscript, so that they could revise their writing according to the principles of coherence mentioned previously. Figure 2 gives an example of the kind of revision students could make on the basis of the analysis of the graphs extracted by the text mining tool. In the example, the text in question was a research project written by a student who was interested in using images and other information to retrieve works of art of potential interest from the internet. The mining of the complete text resulted in a graph with terms such as *Art, student, information, image, works [of art]* (left side of Figure 2). When only the Introduction section of the text was mined, a smaller graph was produced (right side of the image), with terms that also appeared in the complete graph. Many terms did not appear in the smaller one: *information, student, research, aesthetics, teaching* (terms highlighted with a double box frame on the left-side of Figure 2). Hence, to help the student reflect about the ***continuity*** aspect, he/she was asked to verify whether these terms should have been

brought to light in the Introduction as well, where they did not seem to have been emphasized.

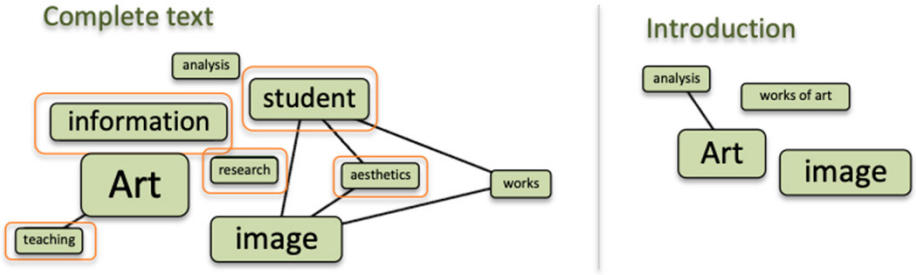

Fig. 2. Graphs obtained from a text about the use of technology to support art learning

The same type of revision protocol was proposed for every section of the text (*Introduction, Background Knowledge, Methods*). It is important to notice that, by highlighting relevant terms of each section and getting students to think about how and where they were presented, students were led to review their projects from a macro-level perspective.

To help the same student revise his/her project in terms of ***progression***, graphs were produced for each section of the text, and the student was asked to observe how terms progressed from one section to another. Figure 2 shows the graphs obtained from the same Art related project and how terms that appeared in an initial graph also appeared in subsequent ones.

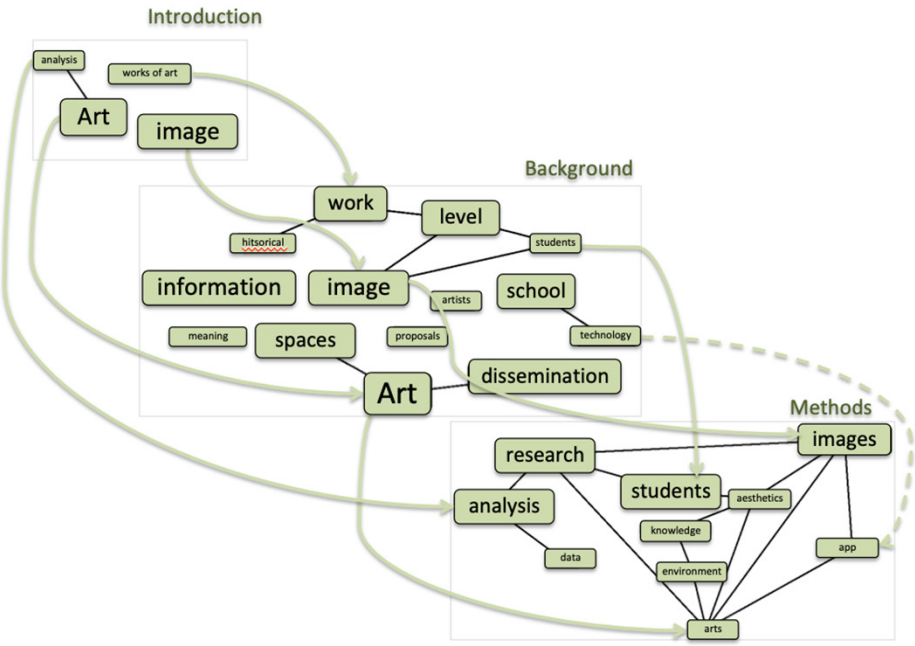

Fig. 3. Progression of terms obtained from subsequent parts of a text about art learning

The term Art appeared in the graphs referring to all sections of the document, which is natural as this was the central topic of the project. Arrows in the figure show this recurrence from one part of the text to another. The word analysis can be spotted in the initial section, it disappears in the Background knowledge section, but it then reappears in the section dedicated to the Methods, which also seems to be a natural progression. However, the term "works of art" that appeared in the Introduction and Background sections, disappeared in the Methods. This may have been an intentional or unintentional decision. Asking the students to think about the recurrence or disappearance of terms should make them more aware of the progression of ideas expressed in the text. The relationship of these ideas with new key elements is also essential, considering ***congruency*** aspects that are part of the evaluation framework used in this research. In this context, connections represented in the graphs could be examined in order to find relationships between ideas. The students were asked to revise these relationships and verify whether they were relevant. Another question is if there were missing connections in the graph, and whether these relationships were stressed in the text with the intended intensity.

As for the factor of ***non-contradiction***, students were asked to verify whether connected terms in the graph were translated into ideas that were not contradictory in different portions of the text. This analysis was not particularly supported by the visualization of the graphs, the question was kept in the analysis model so that it would encompass all of the principles defined by Charolles' framework [17].

5 Results

The essays of the students that were part of the experimental and control groups were marked by 6 different lecturers, who participated in a training course to get familiar with the specific correction grid used in the experiment, in which the principles of continuity, congruence, progression and non-contradiction were emphasized. The lecturers received each essay identified with only a number, so that they would not know whether it belonged to a student that was part of the control or the experimental group. Table 1 shows the results obtained by the students of both groups.

Table 1. Average grades of the students of the control and experimental groups

	N	Average Grade	Standard Deviation
Control group	211	5.284	1.224
Experimental group	285	6.930	1.421

The average grade of the students of the experimental group was higher than that of the control group, 6.930 and 5.284 respectively. The t-Test was then used to verify the statistical significance of the difference between the two averages, resulting in a p-value of 0.002. As this value was smaller than the significance level of 0.05 established for the test, it was possible to assert that the experimental group had a superior average grade than the control group.

In order to better understand how the tool may have helped students in the revision task, learners were asked to leave free comments about the use of the tool.

To analyze these comments from a qualitative perspective, a content analysis technique proposed by Bardin [42] was used. The technique considers the presence or absence of features in a given message fragment. A priori codes and categories of analysis were established prior to examining the data, according to the classes of support the tool provides to students: ***Continuity*** (meaningful repetition); ***Congruency*** (clarity of associations); ***Progression*** (association and conceptual development); ***Non-Contradiction*** (lack of consistency). Two additional categories were defined to account for the ***Appreciation*** or ***Criticism*** of the proposed text writing activity using *TexRev*.

One of the most frequent comments made by the students was that the tool helped them think about the way they write. This was an interesting result from a meta cognitive perspective, as the tool drew the students' attention to certain aspects of their text construction, making them aware of potential macro-level problems in their writings. Below a few excerpts of the students' comments are presented together with their associated categories of analysis.

- *"The tool made me analyze critically the way I write. It is valuable to help in the production of academic texts, it makes us think about the way we organize ideas into words, and the result is the actual text."* ***[Appreciation, Progression]***
- *"The tool qualified my writing in terms of the organization of ideas."* ***[Appreciation, Progression]***
- *"The tool really helped me reformulate part of the final considerations of my text before sending it to evaluation. I also found the presentation of keywords in a mind map interesting to check for coherence problems."* ***[Appreciation, Congruency, Continuity]***
- *"I had never analyzed a text of my own authorship […] I feel more confident now that, even with many elements to be corrected in my text, its quality is adequate."* ***[Appreciation]***
- *"Very good, especially for people like me, with many doubts about writing […] usually I don't like the final result and I end up leaving [the text] as it is when I have to hand it in. But with the use of the tool, it helped me structure the text in a more coherent way."* ***[Appreciation, Progression]***

The category ***Progression*** was the most frequent one, among the ones in the framework for text coherence, showing the students' recognition for how the tool helped them revise their writing in terms of organization of ideas. The category ***Non-contradiction*** was not found in the students comments, which is not surprising as *TexRev* only presented a question asking students to look for inconsistencies in their texts, without providing a specific function to help them find such problems. The category ***Appreciation*** was also present in many comments, while the category ***Criticism*** appeared in a few comments related mostly to usability problems of the tool. The students mentioned "coherence" several times because this was the term used in the writing instructions given by the tool *TexRev* in the proposed activity, following Charolles' text analysis framework [17]. The focus on coherence as a pedagogical strategy brings students to concentrate on aspects of text production that are in the level of discourse, diverting

their aim from specific grammar problems [43], which sometimes is their major and exclusive concern.

6 Discussion and conclusion

In this study we showed how a text mining technique capable of extracting visual representation of texts can be used to help students analyze their own text productions. The study carried out showed that students who used the mining tool achieved higher grades than the students who did not use it.

As for the influence of previous knowledge of graphs on overall results, Yang [44] showed that graph familiarity had no significant impact on writers' performance in a study with 234 students. For this reason, in the experiment presented here it was not a concern whether learners' had previous knowledge of graphs. Still, students of the experimental group were introduced to the text mining tool and its graphical representation of terms before the actual experiment. Results showed that the use of *TexRev* to support writing and revision helped students improve their grades. Whether the students perfected their writing skills and would be able to sustain these results without the use of the text mining tool, this should be a topic for further investigation. In turn, it was demonstrated that the tool was able to provide one-to-one support for the revision of certain text facets at the macro level. The tool's potential to provide support for a large number of students is an important feature, as scaling becomes a desirable aim for the development of tools to support writing and instruction [7]. In this context, the most relevant implication of this research is that the proposed tool to guide students in the revision of their texts following a framework for coherence analysis has the potential to help learners improve their writing from a macro level perspective, with little intervention of teachers.

Comparing *TexRev* with previous work, OpenEssayist [39] has the closest approach to that we followed with the use of text mining to support academic writing. However, although *TexRev* and OpenEssayist are based on the use of a feedback mechanism to get students to reflect about their own writing, the algorithms behind each system are different both in the way text scrutiny is carried out as in the type of feedback provided to students. Previous work on concept mapping has also shown how this strategy can help students improve academic writing, as in [45]. In the work presented in this article, however, the use of graphical representations of texts has been tailored towards diagnostic tasks, and not so much as a planning strategy used in earlier phases of the writing continuum.

As for limitations of this study, *TexRev* is currently not able to provide the means for a deeper analysis of the students' texts. For instance, problems such as leads (e.g. hints of forthcoming sections), transitions (e.g. clarifying how one topic connects to another), and ending (e.g. providing final remarks and explaining in which condition claims made hold true) are hard to identify and correct. Natural language processing features would be probably needed to complement the tool's actual text mining functions if we were to provide support for these kinds of problems. These more complex analyses are possible topics for future work.

7 Acknowledgement

The research presented in this publication has been carried out with the support of CAPES/PROEX, Brazil, and also from the National Agency for Research and Innovation (ANII) in Uruguay, under the code FSED_2_2021_1_169701.

8 References

[1] T. P. Tate, M. Warschauer, and J. Abedi, "The effects of prior computer use on computer-based writing: The 2011 NAEP writing assessment," *Computers and Education*, vol. 101, pp. 115–131, 2016. https://doi.org/10.1016/j.compedu.2016.06.001

[2] A. N. Applebee, and J. A. Langer, "What is happening in the teaching of writing?", *English Journal*, vol. 98, no. 5, pp. 18–28. Available: https://eric.ed.gov/?id=EJ842316 [Accessed March 23, 2022].

[3] P. Morphy, and S. Graham, "Word processing programs and weaker writers/readers: A meta-analysis of research findings," *Reading and Writing*, vol. 25, no. 3, pp. 641–678, 2012. https://doi.org/10.1007/s11145-010-9292-5

[4] T. Haggis, "Pedagogies for diversity: Retaining critical challenge amidst fears of 'dumbing down,'" *Studies in Higher Education*, vol. 31, no. 5, pp. 521–535, 2006. https://doi.org/10.1080/03075070600922709

[5] S. Graham, and K. R. Harris, "Evidence-based writing practices: A meta-analysis of existing meta-analyses," in *Studies in Writing*, R. F. Redondo, K. Harris, and M. Braaksma, Eds. Leiden: Brill Academic Publishers, 2018. https://doi.org/10.1163/9789004270480_003

[6] D. Scott, J. Ribeiro, A. Burns, P. Danyluk, and S. Bodnaresko, "A review of the literature on academic writing supports and instructional design approaches within blended and online learning environments," Werklund School of Education, University of Calgary, Calgary, April, 2017.

[7] C. Rapp, and P. Kauf, "Scaling academic writing instruction: Evaluation of a scaffolding tool (Thesis Writer)," *International Journal of Artificial Intelligence in Education*, vol. 28, no. 4, pp. 590–615, 2018. https://doi.org/10.1007/s40593-017-0162-z

[8] C. Strobl, E. Ailhaud, K. Benetos, A. Devitt, O. Kruse, A. Proske, and C. Rapp, "Digital support for academic writing: A review of technologies and pedagogies," *Computers and Education*, vol. 131, pp. 33–48, 2019. https://doi.org/10.1016/j.compedu.2018.12.005

[9] C. Williams, and S. Beam, "Technology and writing: Review of research," *Computers and Education*, vol. 128, pp. 227–242, 2019. https://doi.org/10.1016/j.compedu.2018.09.024

[10] R. Negretti, "Metacognition in student academic writing: A longitudinal study of meta cognitive awareness and its relation to task perception, self-regulation, and evaluation of performance," *Written Communication,* vol. 29, no. 2, pp. 142–179, 2012. https://doi.org/10.1177/0741088312438529

[11] K. Hyland, *Disciplinary differences: Language variation in academic discourses*. Berlin: Academic Discourse Across Disciplines, 2006. https://doi.org/10.3726/978-3-0351-0446-2

[12] B. Dalton, and B. E. Smith, "Teachers as designers: Multimodal immersion and strategic reading on the internet," *Research in the Schools*, vol. 19, no. 1, pp. 12–25, 2012.

[13] C. H. Hitchcock, L. Rao, C. C. Chang, and J. W. I, Yuen, "Using multimedia tools and scaffolds," *Rural Special Education Quarterly*, vol. 35, no. 2, pp. 10–23, 2016. https://doi.org/10.1177/875687051603500203

[14] J. E. Dowd, T. Duncan, and J. A. Reynolds, "Concept maps for improved science reasoning and writing: Complexity isn't everything," *CBE Life Sciences Education*, vol. 14, no. 4, pp. 1–6, 2015. https://doi.org/10.1187/cbe.15-06-0138

[15] K. Capretz, B. Ricker, and A. Sasak, "Improving organizational skills through the use of graphic organizers," Saint Xavier University and Skylight Professional Development, Chicago, Research Project ED 473 056, 2003.

[16] K. E. Chang, Y. T. Sung, and S. F. Chen, "Learning through computer-based concept mapping with scaffolding aid," *Journal of Computer Assisted Learning*, vol. 17, no. 1, pp. 21–33, 2001. https://doi.org/10.1111/j.1365-2729.2001.00156.x

[17] M. Charolles, "Introduction aux problèmes de la cohérence des textes," *Langue Française*, vol. 38, no. 1, pp. 7–41, 1978. https://doi.org/10.3406/lfr.1978.6117

[18] D. G. Rohman, "Pre-writing the stage of discovery in the writing process," *College Composition and Communication*, vol. 16, no. 2, 1965. https://doi.org/10.2307/354885

[19] D. McCutchen, "From novice to expert: Implications of language skills and writing-relevant knowledge for memory during the development of writing skill," *Journal of Writing Research*, vol. 3, no. 1, pp. 139–153, 2011. https://doi.org/10.17239/jowr-2011.03.01.3

[20] L. S. Flower, and J. R. Hayes, "A cognitive process theory of writing," *College Composition and Communication*, vol. 32, no. 4, pp. 365–387, 1981. https://doi.org/10.2307/356600

[21] N. A. Chenoweth, and J. R. Hayes, "The inner voice in writing," *Written Communication*, vol. 20, no. 1, pp. 99–118, 2003. https://doi.org/10.1177/0741088303253572

[22] A. Becker, "A review of writing model research based on cognitive processes," in *Revision: History, theory, and practice*, A. Horning and A. Becker, Eds. West Lafayette, Indiana: Parlor Press LLC, pp. 25–49, 2006. https://doi.org/10.1080/09585176.2010.504574

[23] J. R. Hayes, L. S. Flower, K., Schriver, J. F. Stratman, and L. Carey, "Cognitive processes in revision," in *Advances in applied psycholinguistics: Reading, writing, and language processing*, S. Rosemberg, Ed. Cambridge: Cambridge University Press, pp. 176–240, 1987.

[24] J. Jones, "Losing and finding coherence in academic writing," *TESOL*, vol. 2, no. 2, pp. 125–148, 2007.

[25] L. Y. Li, and C. C. Tsai, "Students' patterns of accessing time in a text structure learning system: Relationship to individual characteristics and learning performance," *Educational Technology Research and Development*, vol. 68, pp. 2569–2594, 2020. https://doi.org/10.1007/s11423-020-09780-7

[26] D. Hyerle, *Visual tools for transforming information into knowledge*, New York: SAGE Publications, 2009.

[27] E. Kozminsky, N. Nathan, L. Kozminsky, R. Horowitz, "Concept mapping and writing," In Proc. Concept maps: Theory, methodology, technology, Valleta, Malta, pp. 113– 119, 2012. Available: http://cmc.ihmc.us/cmc2012Proceedings/cmc2012%20-%20Vol%201.pdf [Accessed March 23, 2022].

[28] I. M. R. Al-Shaer, "Employing concept mapping as a pre-writing strategy to help EFL learners better generate argumentative compositions," *International Journal for the Scholarship of Teaching and Learning*, vol. 8, no. 2, 2014. https://doi.org/10.20429/ijsotl.2014.080210

[29] P.-L. Liu, "A study on the use of computerized concept mapping to assist ESL learners' writing," *Computers & Education*, vol. 57, no. 4, pp. 2548–2558, 2011. https://doi.org/10.1016/j.compedu.2011.03.015

[30] R. Ferreira-Mello, M. André, A. Pinheiro, E. Costa, and C. Romero, "Text mining in education. Wires," *Data Mining and Knowledge Discovery*, vol. 9, no. 6, 2019. https://doi.org/10.1002/widm.1332

[31] C. H. Yu, S. A. DiGangi, and A. Jannasch-Pennell, "Using text mining for improving student experience management in higher education," in *Cases on innovations in educational marketing: Transnational and technological strategies*, P. Tripathi and S. Mukerji, Eds. Hershey, PA: IGI Global, pp. 196–213, 2011. https://doi.org/10.4018/978-1-60960-599-5.ch012

[32] Y. Xu, and N. Reynolds, "Using text mining techniques to analyze students' written responses to a teacher leadership dilemma," *International Journal of Computer Theory and Engineering*, vol. 4, no. 4, pp. 575–578, 2012. https://doi.org/10.7763/IJCTE.2012.V4.535

[33] I. D. Wahyono, D. Saryone, H. Putranto, K. Asfani, H. A. Rosyid, M, Sunarti, M. M. Mohamad, M. N. H. B. Said, G. J. Horng, and J.-S., Shih, "Shared nearest neighbour in text mining for classification material in online learning using mobile application," *International Journal of Interactive Mobile Technologies*, vol. 16, no. 4, pp. 159–168, 2022. https://doi.org/10.3991/ijim.v16i04.28991

[34] L. Wang, Y, Huang, M. K. Omar, "Analysis of blended learning model application using text mining method," *International Journal of Emerging Technologies in Learning*, vol. 16, no. 1, pp. 172–187, 2021. https://doi.org/10.3991/ijet.v16i01.19823

[35] J. Burstein, "The e-rater® scoring engine: Automated essay scoring with natural language processing," in Automated essay scoring: A cross-disciplinary perspective, M. D. Shermis, and J. Burstein, Eds. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 113–121, 2003.

[36] S. Crossley, L. K. Allen, E. L. Snow, and D. S. McNamara, "Pssst… textual features… there is more to automatic essay scoring than just you!," in Pro, International conference on learning analytics and knowledge, pp. 203–207, 2015. https://doi.org/10.1145/2723576.2723595

[37] J. L. Hsu, H. W., Chou, and H. H. Chang, "EduMiner: Using text mining for automatic formative assessment," *Expert Systems with Applications*, vol. 38, no. 4, pp. 3431–3439, 2011. https://doi.org/10.1016/j.eswa.2010.08.129

[38] E. Reategui, M. Klemann and M. D. Finco, "Using a text mining tool to support text summarization," in Proc. IEEE international conference on advanced learning technologies, Rome, Italy, pp. 607–609, 2012. https://doi.org/10.1109/ICALT.2012.51

[39] D. Whitelock, A. Twiner, J. T. E. Richardson, D. Field, and S. Pulman, "OpenEssayist: A supply and demand learning analytics tool for drafting academic essays," in Proc. International conference on learning analytics and knowledge, pp. 208–212, 2015. https://doi.org/10.1145/2723576.2723599

[40] E. Reategui, M. Bigolin, M. Carniato, and R. A. dos Santos, "Evaluating the performance of SOBEK text mining keyword extraction algorithm," in machine learning and knowledge extraction, A. Holzinger, P. Kieseberg, A. M. Tjoa, and E. Weippl, Eds. Cham: Springer, 2022. https://doi.org/10.1007/978-3-031-14463-9_15

[41] T. Hall, and N. Strangman, "Graphic Organizers," Wakefield, MA: National Center on Accessing the General Curriculum, 2003. Available: https://www.cast.org/products-services/resources/2003/ncac-graphic-organizers-udl [Accessed March 23, 2022].

[42] L. Bardin, "Análise de conteúdo". São Paulo: Edições 70, 2016.

[43] I. Lee, "Helping students develop coherence in writing," *English Teaching Forum*, vol. 40, no. 3, pp. 32–39, 2002. Available: https://americanenglish.state.gov/files/ae/resource_files/02-40-3-i.pdf [Accessed March 23, 2022].

[44] H. C. Yang, "Describing and interpreting graphs: The relationships between undergraduate writer characteristics and academic graph writing performance," *Assessing Writing*, vol. 28, pp. 28–42, 2016. https://doi.org/10.1016/j.asw.2016.02.002

[45] B. A. Pribadi, "Use of the concept mapping strategy to improve academic writing," in educational technology to improve quality and access on a global scale, K. Persichitte, A. Suparman, and M. Spector, Eds. Berlin: Springer, pp. 273–285, 2017. https://doi.org/10.1007/978-3-319-66227-5_21

9 Authors

Simone de Oliveira, has a PhD degree in Education and Technology from the Federal University of Rio Grande do Sul, Brazil. She worked in Basic Education and Higher Education for 20 years. She also worked as a researcher at Instituto Paulo Montenegro in partnership with IBOPE / UNESCO. Currently founder and CPO of GoMining, a company that provides tools and services for text analysis with the use of Artificial Intelligence. E-mail: simone@gomining.com.br

Eliseo Reategui is an associate professor at the Faculty of Education of the Federal University of Rio Grande do Sul, Brazil. He has a PhD degree in Computer Science from the University of London, England, with emphasis in Artificial Intelligence (AI). In the past, he worked with the uses of AI in the industry of games at Kalisto Entertainment, and also at the company Cognition, both in France. In more recent years, he has been a Fulbright Visiting Scholar at University of California Irvine.

Patrícia S. C. C. Barcellos is a professor at the Faculty of Languages and at the Computer Science in Education Graduate Program of the Federal University of Rio Grande do Sul, Brazil. She has a PhD in Computer Science in Education from the Federal University of Rio Grande do Sul and a PhD in Applied Linguistics from the University of Vale do Rio dos Sinos. Her main research interests involve foreign language learning and new technologies applied to education. E-mail: patricia.campelo@ufrgs.br

Marcio Bigolin is a professor at the Federal Institute of Education, Science and Technology of Rio Grande do Sul (IFRS), Campus Canoas, Brazil. He holds a master's degree in Computer Science from the Federal University of Rio Grande do Sul. He is a PhD candidate in Educational Informatics at Federal University of Rio Grande do Sul. His main research interests involve the use of text mining and text review in Education. E-mail: marcio.bigolin@canoas.ifrs.edu.br

Michel Carniato has a bachelor's degree in Electrical Engineering. He has experience in data analysis especially in the development of experimental studies involving text mining research. E-mail: michelcarniato@hotmail.com

Article submitted 2022-03-28. Resubmitted 2022-07-18. Final acceptance 2022-07-18. Final version published as submitted by the authors.