



Associated conference: Shaping the Future of Education in the age of AI: Empowering inclusion, innovation and ethical growth

Conference location: Bologna, Italy

Conference date: 15–17 June, 2025

How to cite: Corino, E., Fissore, C., Conte, M. M. 2025. Data-Driven Methodologies in a Digital Learning Environment for Mathematics. *Ubiquity Proceedings*, 6(1): 20. DOI: <https://doi.org/10.5334/uproc.188>

Published on: 03 September 2025

Copyright: © 2025 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

UBIQUITY PROCEEDINGS

]u[ubiquity press
open access

<https://ubiquityproceedings.com>

Data-Driven Methodologies in a Digital Learning Environment for Mathematics

Elisa Corino¹, Cecilia Fissore², Marina Marchisio Conte²

¹ Department of Foreign Languages, Literatures and Modern Culture, University of Turin, Turin, Italy; elisa.corino@unito.it

² Department of Molecular Biotechnology and Health Sciences, University of Turin, Turin, Italy; cecilia.fissore, marina.marchisio@unito.it

Correspondence: Cecilia Fissore: cecilia.fissore@unito.it

Abstract:

Mathematics is undoubtedly one of the school disciplines in which students generally have the most significant difficulties. Numerous studies in Mathematics education have shown that the causes of disciplinary learning difficulties are the acquisition, understanding, and management of language for specific purposes. This paper will illustrate Data-Driven learning activities with automatic formative assessment implemented within a Digital Learning Environment. The research sample comprises 40 students from two upper secondary schools in Turin, Italy. In an L1 context, students were guided to study the language of Mathematics using a concordance to analyse language and specialised texts. Students carried out activities with automatic formative assessment, featuring step-by-step and adaptive questions that guided them in discovering the peculiarities of the language in use and understanding the concepts expressed by the LPS. In this contribution, we will illustrate this close relationship between the use of technologies, the analysis of language, and the subject contents that the LSP conveys.

Keywords: Automatic formative assessment; Data-Driven Learning; Digital Learning Environment; Mathematics; Mathematics Education; Secondary school.

Introduction

Technologies such as mobile devices, learning environments and tools for language analysis, assessment and translation can play a significant role in language learning. Scientific evidence has shown that using digital applications to support different aspects of learning and teaching a second language positively impacts motivation, proactivity, and longer exposure time. All these aspects favour the long-term assimilation of the study object (Read et al., 2018). This is also true for learning LSP and can be particularly important for disciplines often complex and allegedly incomprehensible to students, such as Mathematics. Numerous studies have demonstrated that language acquisition, comprehension, and management are among the factors contributing to disciplinary struggles (Ferrari, 2004). Two methodologies that exploit the potential of technologies and put the student at the centre of learning are Data-Driven Learning (DDL) and formative assessment. In teaching, DDL is an approach that utilises the tools and techniques of corpus linguistics for pedagogical purposes (Gilquin & Granger, 2010). Introducing a corpus-based approach can support reflection on the language of Mathematics to collect meaningful and realistic data on language use (Cobb & Boulton, 2015) and to reflect on the disciplinary content and its applications. Formative assessment is a process in which students have the opportunity to understand what they have learned and how they have taught it (Black & Wiliam, 2009); therefore, it is one of the teaching practices in which technology can provide fundamental support. In a Digital Learning Environment (DLE) integrated with an automatic assessment system (AAS), it is possible to design and implement activities that incorporate automatic formative assessment, providing immediate and interactive feedback (Corino et al., 2022a). Our research aims to design, implement, and test DDL activities with automatic formative assessment in a DLE, enabling students to develop linguistic skills for analysing and describing mathematical objects and operations. Activities have been designed with personalised, immediate and interactive feedback, providing information on how the DDL task was performed and how to answer the question. Our research questions are:

- (RQ1) What were students' main difficulties in carrying out corpus-based activities with automatic formative assessment?
- (RQ2) Do students appreciate data-driven methodologies?

Literature Review

The importance of linguistic skills in Mathematics

Secondary school students are often required to deal with the languages of the disciplines without having an adequate linguistic framework that enables them to make the necessary distinctions between specialised languages and more general linguistic varieties. Numerous studies in mathematics education have shown that the causes of disciplinary learning difficulties lie in the acquisition, understanding and management of its LSP (Ferrari, 2004). Over time, Mathematics has developed its language with almost unique characteristics of universality, precision, conciseness, and effectiveness (Sbaragli et al., 2021). These characteristics can be complex, distant and restrictive regarding learning. According to Halliday (2004), students' difficulties lie more in grammar than in vocabulary, although the two are inextricably linked. The author states: "To understand why scientific writing became difficult in specific ways, we must eliminate our obsession with words. The difficulty lies more with the grammar than with the vocabulary" (Halliday, 2004, p. 161). The Italian school system is characterised by a clear separation between linguistic and scientific education, which is probably why many students do not apply the linguistic skills they demonstrate in other communicative situations in a mathematical context (Viale, 2019). The acquisition of adequate language skills plays a strategic and essential role, which is not the responsibility of the Italian language teacher alone. However, it is a task to be shared among all the teachers in a class. The complexity of language teaching necessitates collaboration among teachers from different disciplines to ensure that all students are adequately integrated into the school environment and learning pathways. Recently, this aspect has become even more important due to the growing need to teach Mathematics in multilingual class communities composed of individuals with heterogeneous linguistic backgrounds and skills (Ferrari, 2004).

Data-Driven Learning and Automatic Formative Assessment in a Digital Learning Environment

DDL is a didactic approach which treats language as data and sees students as researchers engaged in guided discovery activities (Johns, 1991). The basic idea is that learners can discover grammatical patterns, word meanings or other aspects of language by searching linguistic data and examining large amounts of authentic language. A corpus is a collection of texts, or parts of them, in a finite number, in an electronic format that has been processed uniformly to make it computer-manageable and searchable (Barbera, 2013). Tim Johns (Johns, 1991) argues that using the machine as a rather special kind of informant is the heart of the approach. Once the informant has answered the question, students must attempt to make sense of that answer and integrate it with what is already known. DDL is a methodology particularly suitable for teaching and learning LSP, especially regarding typical word choices (sorted by frequency), nuances of meaning, and the appropriate use of collocations. Students' consultation with corpora can make them more active in forming their knowledge, stimulate higher cognitive processes, and encourage autonomy (Sah, 2015). Furthermore, the procedures, query modes, and prompting skills required by DDL tackle the need for cross-functional competencies related to the daily use of web search tools. Many studies (e.g., Boulton, 2019; Crosthwaite, 2020) reach positive conclusions about the impact of using corpora on language learning. However, DDL work in applied linguistics remains relatively unknown in educational circles (Boulton, 2019; Boulton & Vyatkina, 2021), and it often goes unnoticed – if not unknown – in L1 contexts. We define automatic formative assessment as using formative assessment in a digital learning environment (DLE) by automatically processing student responses and providing feedback (Barana et al., 2021a). One of the five key strategies for implementing formative assessment theorised by Black and Wiliam (2009) is to "provide feedback that moves learners forward". According to Hattie and Timperley, feedback is "information provided by an agent about aspects of one's performance or understanding" (Hattie & Timperley, 2007). A DLE integrated with an automated assessment system can enhance automatic formative assessment and provide immediate, interactive feedback (Barana et al., 2020). Our research group has successfully developed and tested an automatic formative assessment model with interactive and immediate feedback in a DLE (Barana et al., 2021b). We conceptualise interactive feedback as a step-by-step process that guides the learner in solving a task after one or more autonomous attempts (Barana et al., 2021a).

Research context and methodology

The research activity involved four classes from two Italian secondary schools, comprising 80 students in grade 11 and their teachers. The two schools are scientific high schools. The didactic experimentation, from

November to December 2021, consisted of four two-hour meetings in the classroom. The activities involved two classes (40 students) and were paired with two other classes, which formed the control group (40 students). The topic chosen for the experiment was the concept of mathematical function, which was agreed upon by the teachers. For the activities, a specific corpus was created by selecting texts related to the concept of function from secondary school textbooks. The research project combined DDL with formative assessment in a DLE. The DLE was implemented on a Moodle platform (available at <https://linguaggispecialistici.i-learn.unito.it>) integrated with the Möbius Assessment system (<https://www.digitaled.com/products/assessment>). All activities involved the use of two tools in parallel: AntConc (<https://www.laurenceanthony.net/software/antconc/>) for linguistic research and the DLE for carrying out tests with automatic formative assessment and other activities. More than 20 corpus-based activities with automatic formative assessment were designed and implemented within the DLE (Corino et al., 2022b). Then, students carried them out independently under the supervision of trainers. At the end of each three-question set, a group discussion was held to reflect on the emerging concepts and consolidate what had been learnt. To study the effectiveness of these methodologies, students took an initial test at the beginning of the experiment and a final test at the end to evaluate students' mathematical and linguistic competences. These results were compared with the results of the initial and final control group tests. In this analysis, we considered the students' performance in 20 corpus-based activities with automatic formative assessment divided into six tests. To answer the first research question, we observed the students in the two experimental classes as they carried out corpus-based activities with automatic formative assessment to understand the possible difficulties (from both linguistic, disciplinary, and technical perspectives) they might have encountered in using the proposed tools. About the second goal of the research, we analysed the responses of 36 students who participated in the research activity to a final questionnaire. In particular, we analysed a Likert scale question related to satisfaction with the proposed activities and methods (DDL and automatic formative assessment). In this question, students were asked to indicate their level of agreement with different statements on a Likert scale, ranging from "1" (strongly disagree) to "4" (strongly agree). Students were asked whether learning new tools was easy, valuable, and engaging and whether they appreciated the immediate and interactive feedback, finding it compelling. An example of a corpus-based activity with automatic formative assessment is illustrated, along with an example of formative and non-formative feedback related to the activity.

Results

When carrying out the first activities (the first two tests), the students encountered some technical difficulties related to using the AntConc software and the AAS, which they were unfamiliar with. The most common difficulties were:

- Setting up the linguistic search and choosing what to search for in the corpus. Some students stared at the screen, unsure how to proceed, or searched extensively without an explicit criterion. In some cases, when they had difficulty understanding the requests from a mathematical perspective and did not know how to answer a question, they searched for the answer in the corpus, even though they were unsure of what to look for, and proceeded by trial. For example, one of the first questions asked students to select from the list of proposed verbs that function as subjects. Sometimes, students entered the verbs proposed in the answer options in the search bar instead of following the instructions. However, they did not conjugate the verb and therefore did not find results.
- Some students confused the AntConc search bar with that of a search engine (e.g., Google), and they entered combinations of multiple words without finding results. They did not understand that the search was for text strings. For example, they did not understand that "funzione" is different from "funzioni" and that entering an entire phrase made it challenging to find a result, that is, an exactly equal string in the corpus.
- Analysing a large number of results. Most linguistic searches produced many results, and the students tended to analyse only the first ones and respond incompletely to the activities.
- Do not consult the corpus to confirm the meanings of terms. To answer the multipart questions, students did not return to the corpus if something was unclear. Continuing the question in the previous example, the second part asked students to insert the identified verbs (which functioned as subjects) into some sentences. Sometimes, students insert the verbs by trial and error rather than consulting the corpus to clarify the verb's meaning.
- Using the available attempts. Many students did not click the "check" button at the end of the first part of a question. In this way, they not only did not receive immediate feedback on the correctness of their answer and lost the opportunity to correct themselves in case of error by making further attempts, but they also did not see the subsequent parts of the question itself.

All these difficulties were gradually overcome during the project. To this end, it was essential to summarise the concepts that had emerged in the collective discussions at the end of each activity and carry out the test together. Regarding the impact of formative assessment on students' performance, the system's ability to provide accurate data on the performance of activities enabled a thorough analysis of the differences between formative and non-formative approaches. Figure 1 illustrates a corpus-based activity with automatic formative assessment. In the first part, students define the "image of an element" in a function, reflecting on terms and relationships between elements and their images. In the second part, students use AntConc software to search for the word "image," analyse the results, and refine the definition accordingly. They must distinguish between the "image" of an element and the "image set" of a function. The third part requires students to apply the same preposition as in the second part to link elements of two sets. After each part, students can click a "verify" button to receive immediate feedback on their answers, allowing them to retry if needed. This feedback helps them improve their understanding of subsequent parts of the activity.

First part

Completa correttamente:

Se $f: A \rightarrow B$ è una funzione, l'insieme degli elementi dell'insieme B che sono di almeno un elemento dell'insieme A si chiama insieme immagine o semplicemente immagine di A f e si con $Im(f)$ o anche con $f(A)$.

Sezione Tentativo 1 di 3

Verifica

(Clicca per l'Elenco)

- simboleggia
- definisce
- rappresenta
- indica

Second part

Cerca la parola "immagine" inserendo come livelli di concordanza 1L, 2R e 3R.

Completa correttamente:

L'insieme immagine di una funzione è l'insieme degli elementi del che sono immagine di almeno un elemento del , la funzione.

Sezione Tentativo 1 di 3

Verifica

(Clicca per l'Elenco)

- codominio
- dominio

Figure 1: Example of a corpus-based activity with automatic formative assessment.

Figure 2 compares non-formative and formative feedback using two student responses to the question in Figure 1. In the first case (nr1), the student gives incorrect answers for the first two items, using "images" instead of "image" and the wrong preposition "in." Despite three attempts, the student does not change any answers and receives the correct solutions. This feedback is ineffective as the student repeats the same errors. In the second case (nr2), the student initially leaves the second item blank but uses the available attempts and system feedback to revise their answer. The student actively engages with the feedback, utilising the corpus to verify terms and usage, demonstrating a deeper focus on understanding the content rather than just the result. This process provides scaffolding and additional guidance to help the student answer correctly. Table 1 presents students' considerations for using AntConc in linguistic research. They appreciated AntConc for language research because it enabled them to develop new digital and language skills. According to the students, using AntConc (3.09) was easy, understanding how to use it to answer questions (3.14) was also easy, and learning how to use AntConc (3.40) was interesting. They also believe that learning to use text analysis software will be helpful in the future (3.14). The aspects they preferred the most were those concerning automatic formative assessment. Table 2 presents the students' reflections on using the AAS and the potential for receiving immediate and interactive feedback. Students appreciated the possibility of viewing the correct answer immediately after answering a question (immediate feedback) (3.63) and the option to retry the exercise in case of a wrong answer (3.40). They agreed that having the assessment immediately after solving an exercise helped them understand how the question should be answered (3.51). The responses were also optimistic regarding the relationship between language study and mathematical content. According to students, the exercises helped them become more aware of the language to use (3.14), and the exercises on language also made me reflect on meanings and content (2.77). For both the control and experimental groups, the p-value (0.00011368) is below the conventional threshold of 0.05, indicating a statistically significant difference between the pre- and post-test scores. This suggests a significant improvement in test scores from the initial to the final assessment. On average, the control group showed a gain of 1.24 points, while the

experimental group exhibited a larger improvement (+ 2.67 points). These findings indicate that the experimental group outperformed the control group.

Conclusions

The complete analysis of the experimental data is still in progress; however, these preliminary results provide sufficient grounds to encourage the increased use of corpora in educational settings. In response to the first research question, the students encountered difficulties using the proposed technologies to carry out corpus-based activities with automatic formative assessment, including setting up the linguistic search, analysing numerous results, and utilising the available attempts.

Figure 3: Examples of non-formative and formative feedback.

Table 1: Students' thoughts on the use of AntConc for linguistic research

Item	Mean	Std. dev.
It was easy to use AntConc.	3.09	0.85
It was easy to understand how to use AntConc to answer questions.	3.14	0.85
It was interesting to learn how to use AntConc.	3.40	0.74
Learning to use text analysis software is helpful for the future.	3.14	0.91

Table 2. Students' thoughts on the automatic formative assessment.

Item	Mean	Std. dev.
It is helpful to view the correct answer immediately after answering a question.	3.63	0.55
When I gave the wrong answer, I would try the exercise again.	3.40	0.69
The assessment immediately after solving an exercise helped me understand how to answer the question.	3.51	0.66

The students quickly overcame all these difficulties, thanks to the trainers' support and the group discussions. In response to the second research question, students appreciated the proposed methodologies and the mathematical activities, which differed significantly from traditional ones. The DDL approach had a very positive and motivating effect on the students. They also appreciated the automatic formative assessment

method and the opportunity to receive immediate feedback. This allowed them to correct themselves after several attempts and to see the correct answer at the end of each activity. Technology and a student-centred DDL approach enable students to participate in linguistic tasks, particularly in STEM subjects to which they are not accustomed.

References

- Barana, A., Fissore, C., & Marchisio, M. (2020). Automatic Formative Assessment Strategies for the Adaptive Teaching of Mathematics. In *International Conference on Computer Supported Education*, Springer, Cham, 341-365.
- Barana A., Marchisio M. (2021). Analysing Interactions in Automatic Formative Assessment Activities for Mathematics in Digital Learning Environments, *International Conference on Computer Supported Education, CSEDU - Proceedings*, 1, 497–504.
- Barana A., Marchisio M., Sacchet M. (2021a). Effectiveness of Automatic Formative Assessment for learning Mathematics in Higher Education, *International Conference on Higher Education Advances*, 1-8, [10.4995/HEAd21.2021.13030](https://doi.org/10.4995/HEAd21.2021.13030).
- Barana, A., Marchisio, M., Sacchet, M. (2021b). Interactive feedback for learning Mathematics in a digital learning environment", in *Education Sciences*, vol. 11(6), 2021b, 279.
- Barbera, E. F. (2013). *Linguistica dei corpora e linguistica dei corpora italiana. Un'introduzione*. Milano: Qu. ASAR.
- Black, P. & Wiliam, D. (2009). Developing the Theory of Formative Assessment. *Educ. Assess. Eval. Account*, 21, 5–31.
- Boulton, A. (2019). Data-driven learning for younger learners: Obstacles and optimism. In *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners*, Routledge, 14-20.
- Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning and Technology* 25(3), 66–89.
- Cobb, T., Boulton, A. (2015). Classroom applications of corpus analysis. In D. Biber, & R. Reppen, Eds., *Cambridge Handbook of English Corpus Linguistics*, 478–497. doi: [10.1017/CBO9781139764377.027](https://doi.org/10.1017/CBO9781139764377.027)
- Corino, E., Fissore, C., & Marchisio, M. (2022a). Adaptive Exercises and Formative Assessment for English Remedial Action. In *Orchestration of Learning Environments in the Digital World*. Cham: Springer International Publishing, 3-19.
- Corino, E., Fissore, C., & Marchisio, M. (2022b). Data Driven Learning activities within a Digital Learning Environment to study the specialised language of Mathematics. In *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*, IEEE, 167-176.
- Corino, E.; Onesti, C. (2019). Data-Driven Learning: A Scaffolding Methodology for CLIL and LSP Teaching and Learning. *Frontiers in Education*, 4 (7), 1-12.
- Crosthwaite, P. (2020). *Data-Driven Learning for the Next Generation*, New York: Routledge,.
- Ferrari, P. L.. (2004). *Matematica e linguaggio: quadro teorico e idee per la didattica*. Pitagora.
- Gilquin, G., Granger, C. (2010). How can data-driven learning be used in language teaching? In: A. O'Keeffe & M. McCarthy (eds.), *Routledge handbook of corpus linguistics*, London: Routledge, 359-370.
- Halliday, M. A. (2004). *The language of science*. Londra: Continuum.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81-112.
- Johns, T. (1991). Should you be persuaded: two examples of data-driven learning. *Classroom Concordancing: ELR Journal*, 1-12.

- Read, T., Montaner, S., Sedano, B. (2018). Technological Innovation for Specialised Linguistic Domains: Languages for Digital Lives and Cultures. Proceedings of TISLID'18. Mauritius: Éditions universitaires européennes.
- Sah, P. K. (2015). Using Data-Driven Learning (DDL) to Enhance Learner Autonomy. IATEFL SIG Independent, 64, 12-15.
- Sbaragli S., Demartini S. & Franchini E. (2021). Le difficoltà di comprensione e di gestione dei termini specialistici della geometria all'ingresso della scuola secondaria di primo grado, in *La matematica e la sua didattica*, 29(1), 7-37.
- Viale, M. (2019). I fondamenti linguistici delle discipline scientifiche. *L'italiano per la matematica e le scienze a scuola*. Padova: Cleup.