



Data Article

Dataset of mathematics learning and assessment of higher education students using the MathE platform



Beatriz Flamia Azevedo^{a,b,*}, Maria F. Pacheco^{a,b},
Florbela P. Fernandes^{a,b}, Ana I. Pereira^{a,b}

^a Research Centre in Digitalization and Intelligent Robotics (CeDRI), Instituto Politécnico de Bragança, Bragança 5300-253, Portugal

^b Laboratório Associado para a Sustentabilidade e Tecnologia em Regiões de Montanha (SusTEC), Instituto Politécnico de Bragança, Bragança 5300-253, Portugal

ARTICLE INFO

Article history:

Received 7 February 2024

Revised 14 February 2024

Accepted 19 February 2024

Available online 28 February 2024

Dataset link: [Dataset for Assessing Mathematics Learning in Higher Education \(Original data\)](#)

Keywords:

Active learning

Collaborative learning

E-learning

Educational technology

Student engagement

Learning analytics

Adaptive learning

math

ABSTRACT

Higher education institutions are promoting the adoption of innovative methodologies and instructional approaches to engage and promote personalized learning paths to their students. Several strategies based on gamification, artificial intelligence, and data mining are adopted to create an interactive educational setting centred around students. Within this personalized learning environment, there is a notable boost in student engagement and enhanced educational outcomes. The MathE platform, an online educational system introduced in 2019, is specifically crafted to support students tackling difficulties in comprehending higher-education-level mathematics or those aspiring to deepen their understanding of diverse mathematical topics – all at their own pace. The MathE platform provides multiple-choice questions, categorized under topics and subtopics, aligning with the content taught in higher education courses. Accessible to students worldwide, the platform enables them to train their mathematical skills through these resources. When the students log in to the training area of the platform, they choose a topic to study and specify whether they prefer basic or advanced questions. The platform then selects a set of seven multiple-

* Corresponding author at: Research Centre in Digitalization and Intelligent Robotics (CeDRI), Instituto Politécnico de Bragança, Bragança 5300-253, Portugal.

E-mail address: beatrizflamia@ipb.pt (B.F. Azevedo).

choice questions from the available ones under the chosen topic and generates a test for the student. After completing and submitting the test, the answers are recorded and stored on the platform. This paper describes the data stored in the MathE platform, focusing on the 9546 answers to 833 questions, provided by 372 students from 8 countries who use the platform to practice their skills using the questions (and other resources) available on the platform. The information in this paper will help research about active learning tools to support the improvement of future education, especially at higher educational level. Furthermore, these data are valuable for understanding student learning patterns, assessing platform efficacy, gaining a global perspective on mathematics education, and contributing to the advancement of active learning tools for higher education.

© 2024 Published by Elsevier Inc.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Specifications Table

Subject	Data Engineering
Specific subject area	<i>Assessment of higher education students in Math topics.</i>
Data format	Raw
Type of data	Table
Data collection	The dataset was collected online by the MathE platform system. It includes the answers collected by the MathE platform, for 14 topics and 24 subtopics of mathematical contents taught in higher education courses and are available on the MathE platform. Moreover, each question is associated with a set of keywords that describe the question main contents. The answers are described in terms of correct (0) or incorrect (1), and they were provided by 372 students identified by Identification Number (ID). Besides, it is available the information about the country that each students belongs. Thus, in this dataset there are 9546 instances available.
Data source location	The dataset was provided by higher education students from eight countries: Portugal, Lithuania, Italy, Ireland, Romania, Russia, Spain, and Slovenia.
Data accessibility	Repository name: Instituto Politécnico de Bragança – Data Repository Data identification number: https://doi.org/10.34620/dadosipb/PW3OWY Direct URL to access data: https://dados.ipb.pt/dataset.xhtml?persistentId=doi:10.34620/dadosipb/PW3OWY
Related research article	Azevedo, B.F., Pereira, A.I., Fernandes, F.P., Pacheco, M.F.: Mathematics learning and assessment using MathE platform: A case study. Education and Information Technologies (2021). https://doi.org/10.1007/s10639-021-10669-y

1. Value of the Data

- The dataset provides valuable insights into how students from diverse backgrounds engage with and respond to mathematics questions. This information can help educators and researchers understand learning behaviors, preferences, and challenges, as well as supporting students with different levels of proficiency in higher-level Mathematics.
- The data encompasses responses from students from eight countries, offering a global perspective on learning trends and challenges in Mathematics education. This broad view allows for cross-cultural comparisons and identifies patterns that may be specific to certain regions or applicable universally.
- The data shed light on the effectiveness of personalized learning, as students can make choices on topics, question difficulty levels, and pace of study. Understanding how stu-

dents navigate and benefit from personalized learning experiences contributes to the broader conversation on educational technology and tailored learning approaches.

- The information gathered is crucial for investigations into active learning tools. It provides a basis for assessing the impact of such tools on student engagement, knowledge retention, and overall learning outcomes. This, in turn, supports the ongoing improvement of active learning strategies in higher education.
- The dataset allows for the identification of specific challenges that students face in understanding college-level mathematics. By analyzing common misconceptions, areas of difficulty, and patterns of errors, educators can tailor instructional strategies to address these challenges more effectively.

2. Background

Active learning tools encompass various strategies that go beyond passive listening to lectures, engaging students in activities that promote discovery, information processing, and application tasks [5,8]. Case study research has consistently demonstrated that the implementation of active teaching strategies correlates with increased lecturer attendance, heightened engagement, the development of expert attitudes among students towards the subject, and enhanced overall student abilities [5,6,9,10]. It is important, however, to recognize the significance of traditional lecturing, emphasizing that active learning should always be complemented by well-defined content and clear learning objectives [8].

Individuals differ in their profiles of strengths and weaknesses across different types of intelligence, with educational and professional success hinging on the capacity to bolster strengths and offset weaknesses [7]. The challenge arises from the impossibility of teaching effectively to every student's unique learning style. Customized education based on active learning strategies has the potential to revolutionize the way we approach teaching and learning. By recognizing the unique needs and abilities of individual students, we can help to create a more equitable and effective education.

3. Data Description

MathE is a pioneering collaborative e-learning platform that enhances the users' mathematical learning processes in higher education. Its core objective is to cultivate virtual learning and foster knowledge exchange [1].

MathE goals enhance the quality of teaching, learning, and assessment methods in higher education Mathematics content. This platform is a non-commercial tool, entirely free for those interested in improving their knowledge and understanding of Mathematics. MathE has been online since 2019 at <https://mathe.pixel-online.org>. The platform comprises two main sections: the *MathE Library* and *Students Assessment*. The library offers a collection of videos and written materials to support the acquisition of knowledge; in the *Students Assessment* section, there is a set of multi-choice questions to provide the student training assessment to test if a specific topic that he/she enrolled in is already well understood.

Moreover, MathE provides an associated YouTube channel (available at <https://www.youtube.com/@matheproject4778>) with a collection of videos linked to some topic or subtopic available on the platform. For more details about the platform, refer to [1–3].

The dataset [4] is composed of the data collected on the platform system from January 2019 until December 2023, and it is described in terms of 9546 instances and 8 features, as described below:

- **Student ID:** is a numeric identification (ID) assigned to the student at the moment that he/she registers on the platform. The dataset contemplates 372 distinct students.

Table 1
Data per topics and subtopics.

Topic	Number of answers	Subtopic	Number of answers
Linear Algebra	5726	Eigenvectors and Eigenvalues	130
		Linear Systems	420
		Linear Transformation	2127
		Matrices and Determinants	300
		Vector Spaces	2749
Fundamental Mathematics	818	Algebraic Expressions, Equations, and Inequalities	496
		Elementary Geometry	322
Graph Theory	55	Graph Theory	55
Differentiation	579	Derivatives	317
		Partial Differentiation	262
		Integration Techniques	111
Integration	144	Definite Integrals	18
		Double Integrals	15
		Analytic Geometry	358
Analytic Geometry	358	Complex Numbers	592
Complex Numbers	592	Differential Equations	108
Differential Equations	108	Statistic	340
Statistic	340	Domain, Image, and Graphics	107
Real Functions of a Single Variable	164	Limits and Continuity	57
Probability	128	Probability	128
Optimization	182	Linear Optimization	56
		Nonlinear Optimization	126
		Set Theory	42
Set Theory	42	Numerical Methods	310
Numerical Methods	310		

Table 2
Answers per type and level.

Type of answer	Basic level answers	Advanced level answers
Correct	3617	853
Incorrect	4227	849

- **Student Country:** provides information about the country where the student is studying. The dataset comprises students from 8 countries, as also described in Table 3: Portugal, Lithuania, Italy, Ireland, Romania, Russia, Spain, and Slovenia.
- **ID question:** is a numeric value that identifies each question. The dataset is composed of the answers to 833 distinct questions.
- **Type of answer:** it describes the type of answer provided by the student for each question, in terms of correct answer (1) or incorrect answer (0).
- **Question level:** it defines the level of the question answered. It was considered two levels, basic and advanced. This level was identified by the lecture that submit the question.
- **Topic:** it describes in which topic belongs the question answered, among the ones presented in Table 1.
- **Subtopic:** it describes in which subtopic belongs the question answered, among the ones presented in Table 1.
- **Question keywords:** it defines a set of keywords that describe the main contents involved in the question and selected by the lecturer that submits the question.

Table 2 describes the number of obtained answers considering the type of answer, if it is correct or incorrect, and the level of the question, basic or advanced.

The platform is currently being used by a significant number of users. However, some of these students use the platform only to consult the materials, not to answer the questions. Thus,

Table 3
Data per countries.

Country	Number of students	Number of answers
Portugal	245	5495
Ireland	13	300
Italy	50	1358
Lithuania	56	1443
Romania	1	28
Russia	1	107
Spain	3	60
Slovenia	3	755

the dataset [4] contemplates the answers provided by students from 8 countries, as described in Table 3.

Concerning the keywords, the dataset consists of 194 keywords distributed across the topics and subtopics within the platform. These keywords are intended to elucidate about the primary content covered in the questions and assist users in effortlessly locating the materials they wish to explore, according to specific contents.

4. Experimental Design, Materials and Methods

MathE is a collaborative platform maintained by a group of lecturers who voluntarily create, insert, and review questions and materials to contribute to the platform's content. Thus, lecturers are tasked with contributing questions and materials to support the platform. When they wish to add a question, they are required to select a topic, subtopic, and level for the question in the system. Additionally, they must associate written materials or videos that aid in the student learning process, specifically related to the content covered in that particular question. Subsequently, the question and associated materials undergo a review process by a lecturer reviewer, who determines whether the content is approved or rejected for inclusion on the platform. If approved, the question becomes part of the available set for students to attempt, and the associated materials are incorporated into the MathE library. The library can be accessed directly, or the materials may be suggested to students after completing a test. Lecturers have the flexibility to introduce new questions as needed. Consequently, the dataset does not exhibit a balanced distribution of answers per question. In other words, a recently added question typically has fewer answers compared to an older one.

On the other hand, within the *Students Assessment* section, students can improve their knowledge by engaging with the questions provided on the platform. To access these questions, students must specify the topic, subtopic, and difficulty level (basic or advanced) they wish to focus on for their practice. With this information, the platform generates a test comprising seven randomly chosen questions from the MathE data system, aligning with the students' specified criteria. Upon completing the test, the platform provides feedback by indicating which questions were answered correctly and which ones were answered incorrectly. For incorrectly answered questions, the platform suggests associated materials (videos, written notes, solved exercises) to guide the students in moving forward in their learning path. Additionally, the platform records information such as student ID, the questions attempted, and the types of answer in its database.

In this manner, by cross-referencing the information related to both the questions and the students, the dataset presented in [4] is derived. Fig. 1 illustrates the MathE data system.

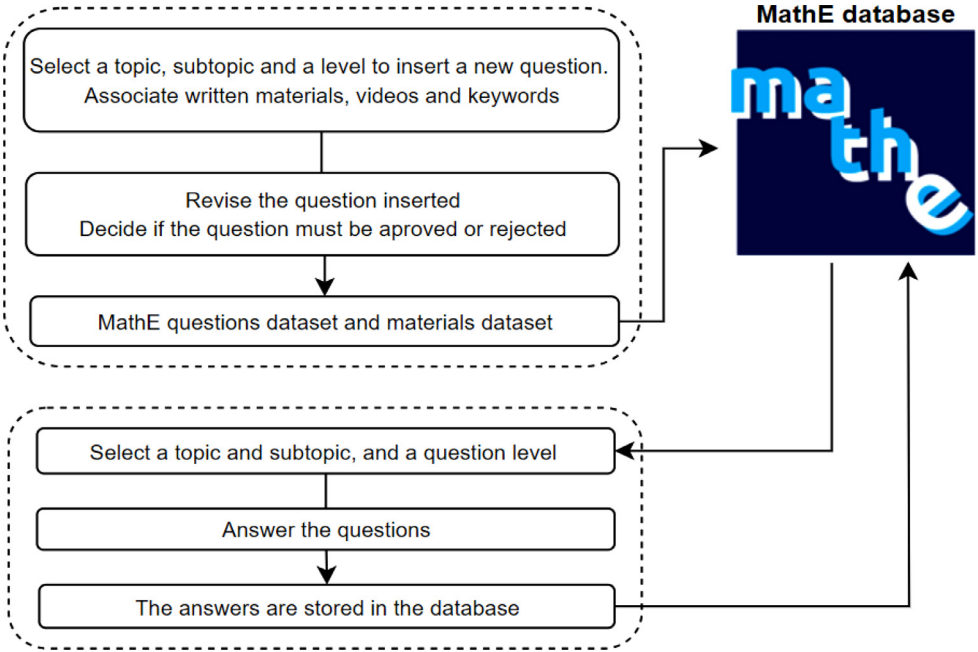


Fig. 1. MathE data system.

Limitations

The items described below do not strictly constitute limitations of the database, but should be considered carefully, depending on the type of analysis one wishes to perform:

- Some questions have been answered many times, while others have been answered few or only once.
- The number of answers to basic questions is much higher than the number of answers to advanced questions.
- The dataset is not balanced in term of country: the majority of answers were provided by students from Portugal, Lithuania, and Italy.

Ethics Statement

The authors have read and follow the ethical requirements for publication in Data in Brief and confirm that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

Data Availability

[Dataset for Assessing Mathematics Learning in Higher Education \(Original data\)](#) (Instituto Politécnico de Bragança – Data Repository).

CRediT Author Statement

Beatriz Flávia Azevedo: Conceptualization, Methodology, Data curation, Software, Investigation, Writing – original draft, Writing – review & editing; **Maria F. Pacheco:** Conceptualiza-

tion, Visualization, Supervision, Investigation, Project administration, Writing – review & editing; **Florbela P. Fernandes:** Conceptualization, Visualization, Supervision, Investigation, Project administration, Writing – review & editing; **Ana I. Pereira:** Conceptualization, Visualization, Supervision, Validation, Investigation, Funding acquisition, Project administration, Writing – review & editing.

Acknowledgments

This work has been supported by FCT Fundação para a Ciência e Tecnologia within the R&D Units Project Scope [UIDB/05757/2020](#), [UIDP/05757/2020](#), [SusTEC LA/P/0007/2021](#) and Erasmus Plus KA2 within the project [2021-1-PT01-KA220-HED-000023288](#). Beatriz Flãmia Azevedo is supported by FCT Grant Reference [SFRH/BD/07427/2021](#).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] B.F. Azevedo, A.I. Pereira, F.P. Fernandes, M.F. Pacheco, Mathematics learning and assessment using MathE platform: a case study, *Educ. Inf. Technologies* (2021), doi:[10.1007/s10639-021-10669-y](#).
- [2] B.F. Azevedo, Y. Amoura, A.M.A.C. Rocha, F.P. Fernandes, M.F. Pacheco, A.I. Pereira, Analyzing the MathE Platform through clustering algorithms, in: O. Gervasi, B. Murgante, S. Misra, A.M.A.C. Rocha, C. Garau (Eds.), *Computational Science and Its Applications – ICCSA 2022 Workshops. ICCSA 2022*, Springer, Cham, 2022 *Lecture Notes in Computer Science*, vol 13378, doi:[10.1007/978-3-031-10562-3_15](#).
- [3] B.F. Azevedo, R.M. de Souza, M.F. Pacheco, F.P. Fernandes, A.I. Pereira, Application of pattern recognition techniques for MathE questions difficulty level definition, in: A.I. Pereira, A. Mendes, F.P. Fernandes, M.F. Pacheco, J.P. Coelho, J. Lima (Eds.), *Optimization, Learning Algorithms and Applications. OL2A 2023. Communications in Computer and Information Science*, Springer, Cham, 2024 vol 1981, doi:[10.1007/978-3-031-53025-8_21](#).
- [4] B.F. Azevedo, F.P. Fernandes, M.F. Pacheco, A.I. Pereira, Dataset for assessing mathematics learning in higher education, *Inst. Polit. Brag.* (2024), doi:[10.34620/dadosipb/PW3OWY](#).
- [5] N.M. Alhawiti, The influence of active learning on the development of learner capabilities in the college of applied medical sciences: mixed-methods study, *Adv. Med. Educ. Pract.* 14 (2023) 87–99, doi:[10.2147/AMEPS392875](#).
- [6] L. Deslauriers, L.S. McCarty, K. Miller, K. Callaghan, G. Kestin, Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom, *Proc. Natl. Acad. Sci.* 116 (39) (2019), doi:[10.1073/pnas.1821936116](#).
- [7] H. Gardner, *Intelligence Reframed: Multiple Intelligences for the 21st Century*, Basic Books, USA, New York, 1999.
- [8] C. Meyers, T.B. Jones, *Promoting Active Learning: Strategies for the College Classroom*, Jossey-Bass Inc., USA, 1993, doi:[10.1016/0307-4412\(94\)90198-8](#).
- [9] X. Gu, Y. Tan, X. Wu, Z. Cai, Q. Lai, M. Cheng, Q. Zhao, Active versus passive strategy in online creativity training: how to best promote creativity of students with different cognitive styles? *Think. Skills Creat.* 44 (2022), doi:[10.1016/j.tsc.2022.101021](#).
- [10] L.S. Keiler, Teachers' roles and identities in student-centered classrooms, *Int. J. STEM Educ.* 5 (34) (2018), doi:[10.1186/s40594-018-0131-6](#).