ACTA SCIENTIAE
ISSN: 2178-7727
DOI: 10.64856/acta.scientiae.8450

Generative Artificial Intelligence and Project-Based Learning Outputs in Technology-Enhanced Mathematics Education

José Manuel Dos Santos Dos Santos [a]
Jaime Carvalho e Silva [b]
Alexandre Manuel Batista Trocado [c]
Zsolt Lavicza [d]

[a] University of Coimbra and Center for Research & Inovation in Education, Portugal
b University of Coimbra and Center of Mathematics of University of Coimbra, Portugal
c Center for Research and Development in Mathematics and Applications, Aveiro, Portugal
d School of Education - Johannes Kepler University, Linz, Austria

ABSTRACT

Background: Background: Generative Artificial Intelligence (GAI) is reshaping technology-enhanced mathematics education, yet little is known about how master's students orchestrate GAI with dynamic and computational tools in project-based learning. **Objectives:** To analyse the mathematical, technological and pedagogical characteristics of products developed by master's-level mathematics education students in a GAI-supported project-based learning sequence. **Design:** Qualitative content analysis combining deductive categories, informed by recent guidance on qualitative and thematic analysis, with inductive, data-driven coding. **Setting and Participants:** A master's programme in mathematics education with five dyads engaged in a three-week technology-enhanced project in which GAI was used. **Data collection and analysis:** The corpus comprised two GeoGebra books and thirteen PDF artefacts (presentations and individual reflections). Data were coded using deductive categories and refined inductively to capture patterns in mathematical reasoning, technological orchestration and task design. **Results:** The study identifies ways in which dyads orchestrated GAI with computational and dynamic geometry tools, and describes patterns of mathematical correctness, justification and classroom-oriented task types. **Conclusions:** The findings indicate the potential of GAI to support task design and work with digital tools, while underscoring the need for critical verification practices and intentional pedagogical integration in project-based learning.

Keywords: Generative Artificial Intelligence; project-based learning; mathematics education; technology integration.

Corresponding author: José Manuel Dos Santos Dos Santos. Email: dossantos@uc.pt

Inteligência Artificial Generativa e Produtos de Aprendizagem Baseada em Projetos no Ensino da Matemática Mediado por Tecnologia

RESUMO

Enquadramento: A Inteligência Artificial Generativa (IAG) está a transformar o ensino da Matemática mediado por tecnologia, mas ainda se sabe pouco sobre a forma como estudantes de mestrado articulam a IAG com ferramentas dinâmicas e computacionais em contextos de aprendizagem baseada em projetos. **Objetivos:** Analisar as características matemáticas, tecnológicas e pedagógicas dos produtos desenvolvidos por estudantes de mestrado em Educação Matemática numa sequência de aprendizagem baseada em projetos apoiada por IAG. **Desenho:** Análise qualitativa de conteúdo, combinando categorias dedutivas, informadas por orientações recentes sobre análise qualitativa e temática, com codificação indutiva baseada nos dados. **Contexto e participantes:** Curso de Mestrado em Educação Matemática com cinco díades envolvidas num projeto de três semanas mediado por tecnologia, no qual foi utilizada IAG. **Recolha e análise de dados:** O corpus integrou dois livros GeoGebra e treze artefactos em PDF (apresentações e reflexões individuais). Os dados foram codificados com recurso a categorias dedutivas e refinados indutivamente para captar padrões no raciocínio matemático, na orquestração tecnológica e na conceção de tarefas. **Resultados:** O estudo identifica formas de orquestração da IAG com ferramentas computacionais e de geometria dinâmica, e descreve padrões de correção matemática, justificação e tipos de tarefas orientadas para a sala de aula. **Conclusões:** Os resultados apontam para o potencial da IAG no apoio à conceção de tarefas e ao trabalho com ferramentas digitais, sublinhando, em simultâneo, a necessidade de práticas críticas de verificação e de uma integração pedagógica intencional em contextos de aprendizagem baseada em projetos.

Palavras-chave: Inteligência Artificial Generativa; aprendizagem baseada em projetos; Educação Matemática; integração de tecnologia.

INTRODUCTION

The integration of digital technologies into mathematics education continues to reshape pedagogical practice, task design, and the orchestration of multiple representations across symbolic, numerical, and visual registers. Contemporary syntheses emphasise the plurality of purposes for technology use, ranging from exploratory modelling to formative assessment, and the need for principled alignment with curricular aims and classroom realities (Drijvers & Sinclair, 2024; Engelbrecht & Borba, 2024; Weigand et al., 2024). Within this landscape, dynamic geometry and computer algebra environments figure prominently as mediational means that can support conceptual understanding and verification-oriented activity when carefully scaffolded (Chytas et al., 2024; Weigand et al., 2024). Reviews of the field underscore both the opportunities and the tensions that arise as new tools are appropriated for teaching and learning, calling for methodologically robust studies that remain sensitive to didactical specificity (Mavrikis & Margeti, 2025; Biehler et al., 2024).

Alongside established environments, classroom use of Generative Artificial Inteligence (GAI) has surged. Literature in higher and secondary education documents potential benefits for feedback, explanation, and task generation, while also foregrounding issues of reliability, academic integrity, and the cultivation of learners' critical agency (Abbas et al., 2024; Bhullar et al., 2024; Cotton et al., 2023; Darvishi et al., 2024). Comparative and design-oriented reports indicate that these assistants can be more productively harnessed when embedded within verification routines and multi-tool workflows, rather than treated as authoritative solvers (Latif & Zhai, 2024; Schorcht et al., 2024; Torres-Peña et al., 2024). Such positioning coheres with wider arguments for purposeful technology use in mathematics education that places teacher mediation and curricular calibration at the centre (Drijvers & Sinclair, 2024; Engelbrecht & Borba, 2024).

The present study examines a bounded corpus of artefacts produced by five dyads enrolled in a master's-level unit focused on computational means for teaching mathematics. Over three weeks, dyads pursued project-based inquiries on algebraic equations of low degree, plane-geometry problem solving, optimisation via extrema, and notable points in triangles, were was used GAI as another technological resource. Two dyads authored GeoGebra books as primary outputs; all dyads produced presentations and individual reflections. Students are referenced solely by codes within dyads (A–E), preserving confidentiality while enabling within- and cross-dyad analysis. The educational setting and the complete list of artefacts are detailed in the course project brief.

This article contributes an in-depth qualitative account of how dyads mobilised dynamic geometry, computer algebra, and GAI to formulate, justify, and verify mathematical work, and how their design choices negotiated classroom feasibility and curricular expectations. It thus addresses current calls to move beyond generalised claims by analysing authentic artefacts and by articulating the forms of mediation and verification that underpin productive technology use in mathematics classrooms (Drijvers & Sinclair, 2024; Engelbrecht & Borba, 2024; Weigand et al., 2024).

METHODS

This section presents the qualitative methodological framework of the study, specifying the reflexive content and thematic analytic stance adopted, the master's-level technology-enhanced mathematics education setting and participating dyads, the corpus of digital and textual artefacts examined, and the procedures, ethical safeguards and quality criteria that structured data generation, coding and interpretation.

Design and analytic stance

The study adopts a qualitative design that combines reflexive content analysis with thematic analysis in a hybrid deductive–inductive configuration. Reflexive content analysis structures descriptive reduction and patterning of meanings across heterogeneous artefacts while maintaining transparency and reflexivity about interpretive decisions (Nicmanis, 2024). Thematic analysis provides a coherent pathway from coding to the development of defensible themes and conceptual contributions, with stepwise guidance appropriate for complex qualitative datasets (Naeem et al., 2023; Zairul, 2025). The hybrid stance is warranted by the dual aims of securing comparability across dyads through a concise set of a priori dimensions and allowing data-driven elaboration of those dimensions through inductive coding (Proudfoot, 2023). Reporting is aligned with values-based qualitative guidance that emphasises methodological coherence and explicit reflexivity (Braun & Clarke, 2025).

Context and corpus

The empirical context is a master's-level curricular unit in Portugal devoted to computational means in mathematics teaching. The project-based sequence spanned three consecutive weeks with one guided session per week and autonomous dyad work between sessions. The dataset comprises fifteen artefacts produced by five dyads (A–E): two GeoGebra Books as principal outputs and thirteen PDF documents containing dyadic presentations and individual reflections. Dyads and artefacts are enumerated in the course brief, which also records that Dyad A and Dyad E published GeoGebra Books linked from the project description.

Data preparation

All artefacts were ingested into a qualitative analysis workspace. Textual content from presentations and reflections was segmented into analytic excerpts at the level of claims, rationales, and design decisions; embedded figures and screenshots were preserved via structured descriptive memos to retain local context. For the two GeoGebra books, navigational structure, construction sequences, and embedded prompts (Used in GAI) were documented through protocolled walk-through notes. A dyad compendium interleaving excerpts and memos was assembled for each dyad to support constant comparison across artefacts and teams.

Deductive coding frame

The deductive frame specified six dimensions derived from the literature and adapted to the corpus: mathematical correctness and justification; pedagogical orchestration with attention to clarity and classroom feasibility; integration of tools across dynamic geometry and computer algebra; critical evaluation and verification practices; alignment with curricular expectations; and the treatment of diagrams and visual reasoning. These dimensions reflect established purposes for digital technology in mathematics education and contemporary attention to assessment and accountability in technology-rich classrooms (Drijvers & Sinclair, 2024; Engelbrecht & Borba, 2024; Weigand et al., 2024). They served as an initial lens for within- and cross-dyad comparison and were open to refinement through inductive analysis (Naeem et al., 2023; Proudfoot, 2023).

Inductive coding and theme development

Inductive coding proceeded iteratively following an immersion–coding–review cycle. Open codes captured patterns insufficiently encompassed by the deductive frame, including staged decomposition of complex problems, triangulation across tools, diagram-dependence and visual fragility, and curricular calibration of methods. Codes were consolidated through constant comparison and organised into candidate themes that were then evaluated for coherence, distinctiveness, and analytic utility across dyads. Theme refinement was recorded through an audit trail of codebook versions and analytic memos, aligning with reflexive content-analytic practice (Naeem et al., 2023; Nicmanis, 2024; Zairul, 2025).

Trustworthiness and reflexivity

Credibility was pursued through triangulation across artefact types within dyads, systematic searches for negative cases, and maintenance of an audit trail documenting analytic decisions. Given the interpretivist stance, coding agreement was established through negotiated discussion rather than calculation of a reliability coefficient, consistent with hybrid inductive–deductive thematic work (Braun & Clarke, 2025; Proudfoot, 2023). Reflexive notes documented the researchers' positioning regarding technology use in mathematics education and potential influences on interpretation.

Ethical considerations

All personal identifiers were removed prior to analysis. Dyads are referred to by dyad names (A–E) and individual students by codes (A1–E2).

Artefacts were analysed for educational research purposes within the course context and stored in a secure repository with restricted access.

RESULTS

The corpus stems from a three-week sequence in the curricular unit Computational Means in Mathematics Teaching in a master's programme for prospective teachers of the third cycle and secondary education in Portugal. The project-based design combined one guided contact session per week with autonomous dyad work. The class worked on a shared overarching question concerning the pedagogic use of contemporary GAI assistants, and computational environments; in week two each dyad addressed a focused sub-question and, in week three, presented outcomes. The dyads were labelled A to E, and students are cited only by codes (A1–E2). Two dyads (A and E) published part of their work in GeoGebra books as their primary artefact, whereas all dyads submitted PDF reflections and presentations documenting processes and results.

The dataset includes, inter alia, the dyadic presentation on polynomial equations and comparative tool analysis (Dyad B), the plane-geometry problem solving and construction analyses (Dyad C), the optimisation tasks involving extrema of functions (Dyad D), and the work on notable points in triangles supported by a GeoGebra book (Dyad E). Dyad A contributed a GeoGebra classroom/book focused on first- and second-degree equations alongside individual reflections discussing verification and representation issues. In all these datasets was used GAI, the dyads experiments used Language Large Models provided by OpenAI, GPT-4o, and Mistral AI, Pixtral Large.

Methodological Approach

The study adopts a rigorous qualitative content analysis oriented by contemporary methodological synthesis. Reflexive content analysis principles informed the descriptive reduction and patterned analysis of manifest meanings in the dyads' textual and visual artefacts, emphasising transparent analytic decision-making and iterative reflexivity (Nicmanis, 2024). In parallel, thematic analysis procedures were used to progress from coding to theme development and interpretation, with attention to conceptual modelling where appropriate (Naeem et al., 2023). Given the evaluative and descriptive aims across heterogeneous materials, a hybrid deductive–inductive strategy was followed to integrate theoretically grounded coding dimensions with data-driven insights (Proudfoot, 2023).

Analytic reporting adheres to values-based qualitative reporting guidance that prioritises clarity of analytic purpose, coherence, and reflexivity in presenting claims (Braun & Clarke, 2025). Throughout the analysis, codes were developed iteratively, applied to the full corpus, and refined as cross-dyad contrasts accumulated. Theme generation proceeded through constant comparison across dyads and artefact types.

Deductive Coding Framework

The deductive framework specified six dimensions; each defined a priori from the methodological literature while allowing local adaptation during coding. The first dimension addressed mathematical correctness and justification, capturing whether solutions and explanations were valid, complete, and aligned with curricular conventions. The second examined pedagogic orchestration, focusing on the clarity of task statements, scaffolding, and classroom feasibility. The third considered tool integration across computational engines and dynamic geometry, observing whether verification and representation were coherently combined. The fourth targeted evidence of critical evaluation, including triangulation of answers and meta-cognitive commentary about reliability. The fifth dimension addressed alignment with curricular expectations and assessment practices. The sixth focused on treatment of diagrams and visual reasoning. These dimensions operationalised the hybrid stance by providing a consistent lens across dyads while remaining sensitive to emergent specifics (Naeem et al., 2023; Nicmanis, 2024; Proudfoot, 2023).

Inductive Theme Development

Inductive analysis yielded a set of cross-cutting themes that recontextualised the deductive dimensions considering the corpus. A prominent theme concerned prompt refocusing and staged decomposition, where dyads reported that complex tasks benefited from guiding assistance step by step rather than wholesale problem delegation. This pattern was especially visible in geometry, where staged decomposition mitigated misinterpretations of constructions and constraints.

A second theme was visual fragility in diagram-dependent tasks. Dyads described systematic difficulties when problems relied on image interpretation or when requesting illustrative outputs, leading to the recommendation to avoid or carefully mediate image-based inputs.

A third theme foregrounded verification through multi-tool triangulation. Dyads recurrently cross-checked results by combining a dialogue-based assistant with computational engines and dynamic geometry visualisation, using agreement or discrepancy as a trigger for further reasoning.

A fourth theme addressed curricular calibration. Dyads noted mismatches between some automated solution methods and national curriculum expectations, particularly the de-emphasis of discriminant-based routines at specific grade levels, which required careful task wording and evaluation criteria.

A fifth theme, volatility and the case for critical autonomy, captured the observation that ostensibly similar prompts sometimes produced divergent expositions, reinforcing the pedagogic need for students to verify claims rather than adopt outputs passively.

Dyad-Specific Synthesis

Dyad A (A1, A2) focused on first- and second-degree equations and designed a GeoGebra classroom to structure student exploration. Their reflections contrasted process-oriented explanations with the exactness and visual affordances of computational engines, highlighting, for example, the didactic value of step-by-step algebraic reasoning alongside the importance of representing complex roots graphically. They reported the need to decode problem statements to exclude inadmissible solutions (such as negative measures) and advocated the use of dynamic geometry for representation when symbolic tools prioritised computation over explanation.

Dyad B (B1, B2) presented a comprehensive comparative analysis across a dialogue-based assistant, an alternative assistant, a computational engine, and a camera-based solver. They curated a GeoGebra classroom to support pupils' reflection on the affordances and constraints of these tools and constructed a comparison framework that foregrounded reliability, graphical representation, classification of equations, and alignment with curricular expectations. Their individual reflections underscored the value of triangulating methods, the necessity of explicit task phrasing to secure curriculum-congruent procedures, and the teacher's role in orchestrating verification.

Dyad C (C1, C2) engaged with plane geometry tasks involving angles in circles, constructions with equidistance constraints, and classical polygonal constructions. They documented recurrent failures when problems were posed through images or required diagram parsing, and they therefore recommended avoiding image-dependent inputs. They reported improved performance when complex tasks were decomposed into sequenced subgoals and when assistance was redirected with focused prompts. The dyad concluded that assistance could contribute to incremental steps but remains brittle with interpretation-heavy tasks, thereby justifying a design that privileges segmentation and human-led interpretation.

Dyad D (D1, D2) analysed problems on extrema for elementary, transcendental, polynomial, and piecewise-defined functions. They compared assistant outputs with dynamic geometry graphing and a computational engine, observing that numerical approximations and occasional misclassifications surfaced in more complex cases. Their reflections showed how graphical inspection and CAS-based exact computation could both corroborate and challenge assistant responses, positioning verification as a designed classroom activity rather than a remedial step.

Dyad E (E1, E2) explored notable points in triangles. Their reflections indicated that while definitional and theoretical answers were generally serviceable, guidance for stepwise constructions within dynamic geometry required careful teacher mediation, and diagram generation was not trustworthy. The dyad created a GeoGebra book to surface frequent points of incompleteness, purposely guiding learners to recognise limitations and to prefer verifiable constructions.

Cross-Dyad Interpretation

Across dyads, the pattern that emerges is a principled division of labour between tools: dialogue-based assistants were leveraged for explanatory discourse and initial structuring, computational engines for exactness and symbolic or numerical reliability, and dynamic geometry for visual validation and exploration. This division supports an instructional design in which pupils articulate conjectures, examine outputs critically, and marshal multiple representations to converge on warranted conclusions. The more diagram-dependent the task, the more the dyads shifted towards human-led segmentation and post-hoc validation, reinforcing the case for scaffolding that explicitly pairs verbal reasoning with visual and algebraic checks. The corpus also indicates that curricular calibration is essential: teachers must frame tasks and evaluate solutions in ways that reflect national expectations regarding methods and representations at specific year levels.

Trustworthiness and Limitations

The analytic process followed hybrid deductive–inductive procedures with explicit documentation of coding decisions and reflexive engagement with the data. The choice of methods and reporting stance aligns with current guidance on methodological coherence and transparency in qualitative work

(Braun & Clarke, 2025; Nicmanis, 2024; Proudfoot, 2023). Limitations include reliance on artefact-based evidence without classroom video, restricted access to the interactive GeoGebra materials beyond metadata and descriptive notes, and the absence of direct pupil artefacts. These constraints were mitigated by triangulating across dyadic presentations and paired reflections and by maintaining an audit trail from codes to excerpts.

DISCUSSION

The analysis of the dyadic corpus suggests that productive technology use in project-based mathematics hinges on the principled orchestration of multiple tools, rigorous verification routines, and deliberate curricular calibration. Across dyads, a consistent division of labour emerged in which GAI text-based assistants supported initial sense-making and explanation, computer-algebra or engine-based environments contributed exact computation and symbolic manipulation, and dynamic geometry enabled visual validation and exploratory refinement. This blended ecology aligns with contemporary accounts that frame digital technologies as mediational means with distinct purposes that must be aligned to instructional intent rather than treated as interchangeable utilities (Drijvers & Sinclair, 2024; Engelbrecht & Borba, 2024). It also coheres with work that foregrounds the role of dynamic and interactive media in connecting representations for teaching, learning, and assessment in mathematics (Chytas et al., 2024; Weigand et al., 2024).

A salient contribution of the corpus is the normalisation of verification as a designed classroom routine rather than a remedial afterthought. Dyads routinely cross-checked outputs among assistants, engine-based solvers, and dynamic geometry, using agreements to warrant claims and discrepancies to trigger further reasoning. This stance resonates with recent discussions of learner agency and academic responsibility in technology-rich settings, which emphasise critical evaluation over passive adoption and argue for assessment designs that make verification visible (Abbas et al., 2024; Cotton et al., 2023; Darvishi et al., 2024). Comparative classroom reports similarly indicate that the reliability of automation improves when tool use is embedded within structured workflows and when teachers make the criteria for acceptable evidence explicit (Schorcht et al., 2024; Torres-Peña et al., 2024). Although automated scoring and feedback solutions are advancing, the present findings caution that their use ought to be subordinated to pedagogical judgement and complemented by opportunities for learners to justify and test claims (Latif & Zhai, 2024). In this respect, the dyads' comparison frameworks functioned as meta-cognitive artefacts that documented why a result should be trusted, not merely whether it was obtained.

The dyads' geometry work surfaces a second, field-relevant nuance: diagram-dependent tasks remain fragile when mediated exclusively through textual or image-based pathways. Reports of misinterpretation, missing constraints, and visually plausible yet mathematically invalid constructions indicate that successful problem solving required staged decomposition and human-led interpretation where GAI was used. This echoes research that positions dynamic geometry as a site for disciplined inquiry in which visual insight is continuously checked against mathematical invariants, and where computational and interactive tools must be carefully orchestrated to avoid over-reliance on surface features (Chytas et al., 2024; Weigand et al., 2024). The implication is not a retreat from visual media, but rather a call to design tasks that explicitly pair visual reasoning with algebraic and numerical checks, and to train learners to articulate what counts as acceptable diagrammatic evidence within a given curricular frame.

Curricular calibration forms a third thread in the discussion. Several dyads observed mismatches between default solution paths suggested by tools and nationally expected methods or representations at stages of schooling. This observation aligns with syntheses that argue for purpose-sensitive deployment of technology in mathematics education, particularly were assessment regimes and curricular documents prescribe not only results but also methods and forms of justification (Drijvers & Sinclair, 2024; Engelbrecht & Borba, 2024; Weigand et al., 2024). The dyads' insistence on specifying acceptable techniques and on requiring representation-rich justifications signals an emergent professional stance in which teachers manage technological affordances to maintain fidelity to curricular aims.

Methodologically, the hybrid analytic strategy adopted for the study—a combination of reflexive content analysis with inductive–deductive thematic work—proved fit for purpose. The deductive frame secured comparability across dyads by foregrounding mathematical correctness and justification, pedagogic orchestration, tool integration, verification practices, curricular alignment, and diagrammatic reasoning. Inductive coding then elaborated these dimensions with corpus-specific patterns such as staged decomposition, triangulation, and visual fragility. This integration mirrors recent methodological guidance that values transparent decision trails, reflexivity, and the movement from coding to conceptually coherent themes in complex qualitative datasets (Braun & Clarke, 2025; Naeem et al., 2023; Nicmanis, 2024;

Proudfoot, 2023; Zairul, 2025). The approach also supported the articulation of case-anchored interpretations without sacrificing cross-case generalisability within the bounded context of the course.

The findings carry implications for teacher education and for project-based design in mathematics. First, programmes should cultivate comparison frameworks as explicit learning products through which novice teachers evaluate and justify tool use, thereby operationalising critical agency in technology-rich classrooms (Mavrikis & Margeti, 2025; Darvishi et al., 2024), where GAI is another technology available. Secondly, task sequences should be engineered to instantiate verification as routine practice, pairing engine-based exactness and dynamic visualisation with text-based explanation so that learners experience the convergence of evidence rather than the primacy of a single source (Chytas et al., 2024; Weigand et al., 2024). Thirdly, assessment criteria ought to name both outcomes and warranted methods, ensuring alignment between classroom practice and curricular expectations (Drijvers & Sinclair, 2024; Engelbrecht & Borba, 2024). Finally, where interactive artefacts such as GeoGebra books serve as primary outputs, reflective documentation should trace construction logic and validation steps so that the evidential status of diagrams is made inspectable.

The study has limitations that qualify transferability. The corpus derives from a single master's cohort and privileges artefact-based evidence over classroom enactment and pupil productions. Interactive materials, while central to two dyads, are difficult to capture exhaustively in static form, which may occlude aspects of learners' exploratory behaviour. These constraints point to the value of multi-modal extensions—combining artefacts, classroom observations, and learner outcomes—to examine how verification routines, staged decomposition, and curricular calibration function under the temporal and social pressures of real classrooms. Notwithstanding these limitations, the present analysis contributes to ongoing efforts to describe, with didactic precision, how technology can be made to serve mathematical reasoning and classroom feasibility rather than the other way round (Mavrikis & Margeti, 2025; Biehler et al., 2024; Weigand et al., 2024).

IMPLICATIONS FOR MATHEMATICS AND TECHNOLOGY EDUCATION

The findings support a pedagogy that explicitly designs for critical verification, encouraging students' participants to contest and corroborate outputs via alternative representations and tools. Staged decomposition appears particularly powerful in geometry and optimisation, where misinterpretations are common. Teacher mediation should ensure that techniques align with curricular expectations and that diagrammatic reasoning is not outsourced but used as a site for disciplined inquiry where GAI are present. For project-based learning, constructing comparison frameworks—as evidenced by the dyads' own meta-analyses—can itself be a productive assessment artefact, documenting how learners justify trust in results and reconcile divergences across tools.

AUTHORS' CONTRIBUTIONS STATEMENTS

JMDSDS and JCS conceived the presented idea. ZL adapted the methodology to this context. JMDSDS and JCS created the models, performed the activities, and collected the data. JMDSDS and AEBT analysed the data. All authors actively participated in the discussion of the results, reviewed and approved the final version of the work.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study will be made available by the corresponding author JMDSDS upon reasonable request and subject to a data-sharing agreement approved by the authors' institution.

ACKNOWLEDGEMENTS

This research was supported in part by the Centre for Research and Innovation in Education (inED) (https://doi.org/10.54499/UIDP/05198/2020), the Center for Mathematics, University of Coimbra (https://doi.org/10.54499/UIDB/00324/2020), the Center for Research and Development in Mathematics and Applications (https://doi.org/10.54499/UID/04106/2025); through the FCT - Fundação para a Ciência e a Tecnologia, I.P..

REFERENCES

Abbas, M., Zaffar, M., Erum, U., Gao, S., Gu, Z., & Liu, J. (2024). Is it harmful or helpful? examining the causes and consequences of chatgpt usage among university students. International Journal of Educational Technology in Higher Education, 21, 44. https://doi.org/10.1186/s41239-024-00444-7

Mavrikis, M., & Margeti, M. (2025). Review of mathematics education in the age of artificial intelligence. *Research in Mathematics Education*, *27*(3), 587–594. https://doi.org/10.1080/14794802.2024.2389418

Bhullar, P. S., Chauhan, R., Bhullar, M. K., Mahajan, R., & Chauhan, A. (2024). Chatgpt in higher education—a synthesis of the literature and a research agenda. Education and Information Technologies. https://doi.org/10.1007/s10639-024-12723-x

Biehler, R., Hochmuth, R., Schreiber, S., et al. (2024). New trends in didactic research in university mathematics education. *ZDM–Mathematics Education*, *56*(8),b1533–1549. https://doi.org/10.1007/s11858-024-01643-2

Braun, V., & Clarke, V. (2025). Reporting guidelines for qualitative research: a values-based approach. *Qualitative Research in Psychology*, *22*(2), 399–438. https://doi.org/10.1080/14780887.2024.2382244

Chytas, C., Drijvers, P., Kempen, R., Schuvindt, M., & Vos, N. (2024). Computational thinking in secondary mathematics education with geogebra: Insights from an intervention in calculus lessons. *Digital Experiences in Mathematics Education*, *10*, 1–22. https://doi.org/10.1007/s40751-024-00141-0

Cotton, D. R. E., Wise, J., Joyner, M., Viera, A., Wolstencroft, P., Bandura, P., Griffiths, R., & Spark, J. (2023). Chatting and cheating: Ensuring academic integrity in the era of chatgpt. *Innovations in Education and Teaching International*. https://doi.org/10.1080/14703297.2023.2190148

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of ai assistance on student agency. *Computers & Education*, *210*, 104967. https://doi.org/10.1016/j.compedu.2023.104967

Drijvers, P., & Sinclair, N. (2024). The role of digital technologies in mathematics education: Purposes and perspectives. *ZDM–Mathematics Education*, *56*(2), 239–248. https://doi.org/10.1007/s11858-023-01535-x

Engelbrecht, J., & Borba, M. C. (2024). Recent developments in using digital technology in mathematics education. *ZDM–Mathematics Education*, *56*(2), 281–292. https://doi.org/10.1007/s11858-023-01530-2

Latif, E., & Zhai, X. (2024). Fine-tuning chatgpt for automatic scoring. *Computers & Education: Artificial Intelligence*, *6*, 100210. https://doi.org/10.1016/j.caeai.2024.100210

Naeem, M., Ozuem, W., Howell, K., & Ranfagni, S. (2023). A step-by-step process of thematic analysis to develop a conceptual model in qualitative research. *International Journal of Qualitative Methods*, *22*, 1–18. https://doi.org/10.1177/16094069231205789

Nicmanis, M. (2024). Reflexive content analysis: An approach to qualitative data analysis, reduction, and description. *International Journal of Qualitative Methods*, *23*. https://doi.org/10.1177/16094069241236603

Proudfoot, K. (2023). Inductive/deductive hybrid thematic analysis in mixed methods research. *Journal of Mixed Methods Research*, *17*(3), 308–326. https://doi.org/10.1177/15586898221126816

Schorcht, S., Buchholtz, N., & Baumanns, L. (2024). Prompt the problem – investigating the mathematics educational quality of ai-supported problem solving by comparing prompt techniques. *Frontiers in Education*, *9*, 1386075. https://doi.org/10.3389/feduc.2024.1386075

Torres-Peña, C. M., Monedero, J., Pedrosa-Jesus, D., Chacón, J. M., Esteban, F. S., Jiménez, Á., & Díaz-Hernández, J. J. (2024). Updating calculus teaching with ai: A classroom experience. *Education Sciences*, *14*(9), 1019. https://doi.org/10.3390/educsci14091019

Weigand, H.-G., Trgalová, J., & Tabach, M. (2024). Mathematics teaching, learning, and assessment in the digital age. *ZDM–Mathematics Education*, *56*(4), 525–541. https://doi.org/10.1007/s11858-024-01612-9

Zairul, M. (2025). Mastering thematic analysis: A step-by-step guide for researchers. *International Journal of Qualitative Methods*, *24*. https://doi.org/10.1177/16094069251384401