Mind the Knowledge Gap: Evaluating AI Tutors' Ability to Detect Mathematical Prior Knowledge and Misconceptions

Xinyue Li, Frank Morley and Rachad Zaki

March 25, 2025

Mind the Knowledge Gap: Evaluating AI Tutors' Ability to Detect Mathematical Prior Knowledge and Misconceptions

Xinyue Li[1[0009-0007-4973-2439]], Frank Morley[2[0009-0009-8200-9730]] and Rachad Zaki[1[0000-0002-8959-9835]]

[1] Cambridge Mathematics, University of Cambridge, Cambridge, UK
[2] Cambridge University Press & Assessment, University of Cambridge, Cambridge, UK
xinyue.li@cambridge.org
frank.morley@cambridge.org
rachad.zaki@cambridge.org

Abstract. Designing effective mathematics Artificial Intelligence (AI) tutors presents a unique challenge in the field of AI in Education (AIED). While it is possible to train an AI model with strong mathematical knowledge, the role of 'AI as a tutor', specifically, taking on a traditional tutor's key role in reviewing students' misconceptions and knowledge gaps, is an under-explored area. This creates a risk that many AI tutoring experiences cannot be adequately tailored to student needs. Consequently, students may develop persistent misconceptions that go unaddressed because AI systems lack the diagnostic capabilities to identify specific gaps in understanding. Despite the growing integration of AI in education, the mathematics education community has yet to adequately address the foundational tasks of defining and developing pedagogically sound technique to test the capabilities of AI models to correctly detect misconceptions as well as knowledge gaps students have or develop. Many AI-powered resources, including AI tutors, are developed by commercial companies designing large language models, whose priorities are often driven more by market trends and hype than by learning sciences (e.g., addressing the needs and characteristics of learners and teachers). In this paper, we envision a training process for the development of a high-quality mathematics AI tutor embedded with a constructivist approach (emphasising that learners actively construct their own knowledge rather than being passively guided by questions), which is an alternative to the commonly adapted Socratic dialogue, to facilitate effective self-directed mathematics learning. The model would focus on the very first part of the tutoring process, which is identifying student misconceptions and knowledge gaps. In this paper, we propose a technique to test the capabilities of an AI model to correctly detect misconceptions students have developed. This is done so that more personalised and effective tutoring can be achieved.

Keywords: AI Tutor, Constructivism, Mathematics Education, Misconceptions, Prior knowledge

1 Introduction

Research in Intelligent Tutoring Systems (ITS) has focused on improving computer-based tutors to make them more adaptable, autonomous, and responsive to the personalised needs of each learner. Some researchers argue that for an intelligent tutor, whether human or artificial, to be effective in supporting learning, it must have three key areas of knowledge: 1) expertise in the subject matter being taught, 2) an understanding of the personalised learner's needs and progress, and 3) knowledge of effective teaching strategies (Conati, 2009). In turn, an understanding of a learner's needs and progress requires an ability to detect gaps in their knowledge and misconceptions. Educational technologies, including AI tutors, are developed as responses to educational needs; it is important that they are developed in the direction of "curriculum → pedagogy → technology, in the sense that teaching starts from curriculum in terms of the aims of the educational process, which are responded to through pedagogy—informed by theories of learning—and then drawing upon the most appropriate technologies" (Taber, 2017, p. 398).

Learning can be understood as a change in behavioural repertoire (Taber, 2009), and teaching can be understood as action taken to bring about specific learning. Many researchers have argued that the most effective learning takes place when learners engage in learning activities in their Zone of Proximal Development (ZPD) (e.g., Barohny, 2019; Li & Zaki, 2024). ZPD is a spatial metaphor which is developed from the Zone of Actual Development (ZAD) – what we have already mastered and can accomplish well without support – and it can be developed into the Zone of Distal Development (ZDD) – what we have no competency in at all and cannot achieve even with guidance or support; these zones shift, as learners can move along these dimensions depending on the development of competence (Taber, 2018; Taber & Li, 2021).

In a formal education context (e.g., schools, classrooms), a professional teacher is expected to be able to identify each learner's learning needs, ensuring that teaching is situated within the learner's ZPD. In informal learning contexts (e.g., at home, in museums), where in most cases a professional teacher is not present to facilitate learning, learners need to identify their own needs in order to select appropriate resources to fulfil their learning needs. This can be a challenging task, even for well-trained teachers, due to the complexity of the nature of learning. Therefore, we should not expect learners to accurately identify their own ZPDs in self-directed learning contexts, even for gifted learners or those with high levels of metacognitive skills. This is where AI tutors have the potential to transform self-directed learning, but only if they are well-designed.

In the context of mathematics education, the growing number of available training datasets containing large collections of mathematics questions, solutions, and mark schemes (e.g., OpenWebMath, ASSISTments, OpenMathInstruct, etc.) makes it relatively easy to train an AI tutor with expertise in the subject matter (e.g., the ability to answer a mathematics question). However, it is possible for humans (and AI) to perform very well at high-level mathematical problems but still struggle with teaching other people about mathematics. Despite the massive potential of machine learning in identifying each learner's prior knowledge, misconceptions, and knowledge gaps, it remains a significant challenge for AI tutor developers to create systems that can effectively facilitate personalised learning. Currently, this is often achieved through learners providing specific prompts like, "*I am a Year 7 student, can you help me with…?*" while the AI tutor responds according to this context based on a general understanding of a Year 7 learner's level (if the AI tutor has been trained to have a general understanding of the different needs of different year groups). Since learners may have idiosyncratic ideas which are not identified in existing research (and therefore not included in the training data), even though these systems might adapt based on input, the effectiveness of such AI tutors in fostering a truly personalised learning experience is still limited.

Teaching is unlikely to be effective if being delivered (whether by a human or AI) in a non-interactive way, that is without considering learners' prior knowledge, pre-existing misconceptions, or how their ideas evolve (or remain unchanged) during instruction. A major obstacle is designing an AI tutor that not only reacts to queries but also proactively diagnoses and responds to nuanced learning needs. For example, an ideal AI tutor would be able to recognise when a learner is struggling with foundational concepts, adjust the level of difficulty dynamically, and offer personalised feedback and alternative explanations tailored to the learner's unique learning needs. Currently, pedagogically based AI tutors tend to be designed with a Socratic dialogue approach, but "the overuse of direct questioning could frustrate learners, especially when neither the AI tutor (i.e., through interrogation and data analysis) nor the student (i.e., through metacognitive awareness) has identified what the student really needs" (Taber & Li, in press). We believe that a constructivist approach emphasising that learners actively construct their own knowledge rather than being passively guided by questions could be used to facilitate a more effective tutoring experience, and the key is to use the AI tutor's rapid data analysis capabilities to identify the student's ZPD.

In this paper, we envision a training process for the development of a high-quality mathematics AI tutor embedded with a constructivist approach; the model would focus on the very first part of the tutoring process, which is identifying student misconceptions and knowledge gaps. We propose a technique to test the capabilities of an AI model to correctly detect misconceptions students have made, so that a more personalised and effective tutoring experience can be achieved.

2 A Constructivist AI Tutor in Mathematics

2.1 The Foundation for Personalised Learning – Diagnostic and Formative Assessment

Personalised learning is an idea that "has the potential to make every young person's learning experience stretching, creative, fun and successful" (Department for Education and Skills, 2004: 3). The most accurate way to identify personalised needs is through assessment. Diagnostic assessments help teachers identify common misconceptions across various topics, and AI technology can further enhance their effectiveness by incorporating a two-level interactive test approach. The first level could consist of multiple-choice questions, with distractors chosen carefully to include common misconceptions, to obtain definite answers, followed by a second-level test item that presents feasible responses based on the specific first-level answer. For instance, the first-level question could be: *"What is the sum of 1/4 and 1/3? (A) 2/7; (B) 7/12; (C) 2/12"*. The goal of this step is to get a definite answer from the learner, and then depending on the learner's answer, they receive a tailored follow-up second-level question. If the learner chooses the correct answer (B), the AI tutor could then ask them to explain the steps they took to find their answer to ensure understanding. If the learner selects option (A), the AI tutor could point out that it appears they added the numerators and denominators directly. It would then explain that a common denominator is needed to add fractions and encourage the learner to try again by asking: *"What is the least common denominator of 4 and 3?"*. This approach helps diagnose misconceptions, allowing for appropriate scaffolding to support learning. This two-level interactive test approach allows for a more personalised learning experience, enabling AI tutors to adapt their output to individual learner responses and provide more targeted resources, teaching strategies, and feedback.

2.2 Mathematical Reasoning

Mathematical reasoning (MR) is a complicated concept; there is currently no shared definition of it within the educational research community (Jeannotte & Kieran, 2017). Lithner (2000) described reasoning as a four-step structure: a problematic situation, a strategic choice, the implementation of the strategy, and a conclusion. Assessing the mathematical reasoning of Large Language Models (LLMs) remains a challenge. Current benchmarks (e.g., GSM8K, MATH) primarily focus on answer/output correctness, often overlooking the reasoning process and the intermediate steps leading to the solution. Additionally, it is difficult to determine whether a correct answer stems from genuine mathematical reasoning skills or simply from the test items being included in the training data. AI tutors must possess strong mathematical reasoning skills before they can effectively 'teach'. However, having these skills alone does not guarantee an effective learning experience for learners. Effective teaching requires more than just solving problems correctly, it also involves understanding where and why learners struggle, identifying misconceptions in their reasoning, and providing targeted feedback to guide them toward conceptual understanding.

Currently, many AI tutoring systems, even those that claim to exhibit high levels of mathematical reasoning, lack the ability to diagnose learners' individualised misconceptions effectively. Without this diagnostic capability, AI tutors risk providing generic explanations that do not directly address learners' misunderstandings. For AI tutors to truly enhance the mathematics tutoring experience, they must go beyond answer correctness and develop the ability to analyse learners' thought processes, recognise patterns of misconceptions, and adapt their explanations accordingly. Therefore, we believe that a constructivist approach should be adopted.

2.3 A Constructivist Approach

The idea of constructivism can be largely credited to Jean Piaget and Lev Vygotsky; while Piaget's work emphasised the internal processes of human development rather than external influences (Piaget, 1971), Vygotsky (1978) introduced the theory of social constructivism, which emphasised the role of social and cultural interactions (e.g., peer interactions, interactions with cognitive tools, etc.) in shaping mental constructs. Expanding on Vygotsky's ideas, Wood, Bruner and Ross (1976) and other educational psychologists further introduced and developed the notion of scaffolding, a strategy in which learners receive temporary support that is gradually withdrawn as they become more independent in their understanding. In particular, Socratic questioning is often associated with the scaffolding theory (Favero et al., 2024).

The Socratic approach is one of the most commonly incorporated teaching strategies in the field of AI-assisted tutoring. In practice, AI begins by asking open-ended questions to help learners articulate their thoughts. It then encourages reflection by guiding learners to analyse and understand their reasoning. Additionally, AI may prompt learners to explain their answers before confirming their correctness. Although many researchers argue that Socratic-style questioning can foster critical thinking and deep learning (Paul & Elder, 2019), its success with human tutors does not necessarily guarantee its effectiveness in AI tutors as AI (at least at this stage) lacks the ability to recognise frustration or provide motivational support like human tutors. As a result, its attempt to mimic the Socratic approach may sometimes lead to frustration and demotivation among learners, particularly when it fails to recognise when a learner is struggling or in need of direct guidance rather than further questioning. For instance, if a learner struggles with factoring quadratic equations, an AI tool using the Socratic method might repeatedly ask: "*What two numbers multiply to give the constant term and add to give the middle coefficient?*". While this may work for some learners, the wording might confuse some, and others may not even know where to begin, leading to frustration and boredom instead of deeper understanding. This limitation highlights the need for alternative approaches that prioritise active learning and learner engagement.

One such alternative is personal constructivism. Personal constructivism (sometimes also referred to as cognitive/pedagogic/psychological constructivism) emphasises the role of the individual learner in constructing meaning, viewing learning as an active process where new knowledge is built upon prior experiences and understandings; therefore, learning is iterative, interpretative and incremental (Taber, 2014). Instead of relying on generic questioning, AI tutors could adopt a more adaptive and personalised approach to ensure that learners remain actively engaged within their ZPD. This could be achieved by providing personalised guidance, adaptive scaffolding, and tailored feedback. For instance, rather than persistently asking: "*What do you notice about the angles in this triangle?*", an AI tutor could enable interactive exploration, allowing learners to manipulate shapes using AI-generated images. As real-time data is collected, the AI tutor could determine when the learner is ready to move beyond their ZPD and progress to more advanced concepts or when the current learning activities are too challenging, requiring a simpler approach. In such cases, the AI tutor could dynamically adjust the difficulty level, offering alternative teaching activities and materials better suited to the learner's current understanding. By integrating elements of personal constructivism, AI tutoring systems could better support learners by providing greater flexibility, personalised learning pathways, and an optimal balance between inquiry-based exploration and direct instruction. To the best of our knowledge, we are unaware of any fully implemented chatbot or AI tutor designed to explicitly leverage the constructivist approach. The key to such an AI tutoring system lies in accurately assessing the learner's prior knowledge, which in turn allows for the identification of their ZPD. To address this, we propose a method for detecting gaps in knowledge in the next section.

3 Method: Detecting Gaps in Knowledge

Using real student data in testing LLMs involves a host of ethical concerns including privacy and consent. Therefore, we would opt to use an LLM (hereafter, 'student LLM') to generate synthetic data of student responses. We would firstly select an assessment of a mathematics topic at the equivalent of UK GCSE level. This assessment would be parsed and converted into a specific format like JavaScript Object Notation (JSON), where the metadata of the question number and question content would be stored. Additionally, the mark scheme for the questions would also be parsed. The student LLM would be instructed to calculate the correct solutions for each mathematics question using the mark scheme, and this would be validated using techniques such as sequence matching (in comparison to the mark scheme) with human oversight. With the ability of the student LLM to produce correct responses established, the student LLM would then be instructed for certain topics to purposely produce an incorrect answer using a 'misconception scheme', which contains common student misconceptions of these topics. These outputs would in turn be validated by sequence matching (in comparison to the misconception scheme) with human oversight. This process would produce the synthetic data of an assessment with correct responses and misconceptions.

With the data in place, the LLM (hereafter, 'tutor LLM') would be tasked with (1) marking the assessment and (2) drawing out misconceptions and knowledge gaps from the marked assessment. The tutor LLM would mark the assessment using the mark scheme, and their marks would be compared to the labelled data which was created as a byproduct of the synthetic data generation. With the assessment marked, the tutor LLM would produce a summary of the questions and topics the student LLM answered correctly and incorrectly, and this would be compared to a ground truth reference dataset and tested for accuracy. This analysis would therefore establish whether the tutor LLM can find and articulate misconceptions and gaps in knowledge, accurately identifying prior student knowledge in a specific domain.

4 Concluding Remarks

Designing a solid and fully reliable AI tutor presents several challenges. We highlighted some in this paper, but many still need to be tackled. For example, when it comes to STEM-related subjects, the gender gap seems to be an unavoidable topic of conversation. Researchers found that in the most recent international assessment, TIMSS 2023, boys in Year 5 and Year 9 at school in England scored significantly higher than girls in mathematics and science (Richardson et al., 2025).

To get a quick idea of whether gender stereotypes are still present in popular AI chatbots as they were when they were first released to the public (Kaplan et al., 2024), we used the question presented in Section 2.1, "*What is the sum of 1/4 and 1/3?*" as a prompt with a commonly used AI chatbot (not primarily designed to be used as an AI tutor), and asked the chatbot to explain it to a girl and a boy. This chatbot used a chocolate bar as an example when explaining to a girl and used a pizza as an example when explaining to a boy. When we asked why they chose a chocolate bar for a girl and a pizza for a boy, they explained, *"I was just trying to use fun and relatable examples, but I see how that could come across as a stereotype. Honestly, both chocolate and pizza (or anything else) work for anyone, regardless of gender! If you'd prefer, I can use the same example for both or even something completely different—like apples, money, or Legos. The goal is just to make the explanation easy to understand"*. Due to the focus of our paper, we only looked at one example, but we believe that any implementation of AI tutors would need to systematically test whether such tutors have gender bias in the methods they take to personalise learning.

School, as a key aspect of social life, reflects various mechanisms found in society, such as gender stereotypes; and teachers are also influenced by gender stereotypes (Kollmayer, Schober & Spiel, 2018). As a result, (human) teachers, despite consistent attempts and best intentions, cannot guarantee that all students are being treated exactly the same without the influence of pre-existing beliefs. AI tutors exhibit similar biases because they are ultimately trained on human data. Therefore, we believe it is important to continue the efforts to ensure that the personalised tutoring experience is based on knowledge gaps and misconceptions, not gender difference.

When designing and developing an AI tutor, several other key elements must be considered. First, creating valid learner profiles is essential (Fonseca & Mora, 2004), as these serve as the foundation for personalised learning experiences. However, this is also a challenge because the curriculum requirements and the skills and concepts that a learner is expected to master vary by grade level, age group, and geographic location. Additionally, data privacy (Ismail & Aloshiand, 2025) and other ethical concerns must be addressed throughout the design and training of AI tutoring systems. Another important consideration is the digital divide (Li & Zaki, 2024), which has been widened by the rapid advancement of AI; to bridge this gap, promoting AI literacy worldwide is crucial (Gonzales, 2024). We believe that future research in AIED should take all of these factors into account.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

Barohny, E. (2019). The zone of proximal development as an overarching concept: A framework for synthesizing Vygotsky's theories. *Educational Philosophy and Theory*, *51*(1), 18–30. https://doi.org/10.1080/00131857.2017.1421941.

Conati, C. (2009). Intelligent tutoring systems: New challenges and directions. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence*. https://www.ijcai.org/Proceedings/09/Papers/012.pdf.

Department for Education and Skills. (2004). *A National Conversation about Personalised Learning.*

Favero, L., Pérez-Ortiz, J. A., Käser, T., & Oliver, N. (2024). Enhancing critical thinking in education by means of a Socratic chatbot. *arXiv preprint arXiv:2409.05511*.

Fonseca, J. M. & Mora, A. D. (2004). An AI based approach to the learners profile estimation. In *Proceedings of CE2004: The 11th ISPE International Conference on Concurrent Engineering: Research and Applications*.

Gonzales, S. (2024). *AI literacy and the new Digital Divide – A Global Call for Action.* UNESCO.

Ismail, I. A., & Aloshi, J. M. R. (2025). Data Privacy in AI-Driven Education: An In-Depth Exploration Into the Data Privacy Concerns and Potential Solutions. In *AI Applications and Strategies in Teacher Education* (pp. 223-252). IGI Global.

Jeannotte, D., & Kieran, C. (2017). A conceptual model of mathematical reasoning for school mathematics. *Educational Studies in Mathematics, 96*(1), 1–16.

Kaplan, D. M., Palitsky, R., Arconada Alvarez, S. J., Pozzo, N. S., Greenleaf, M. N., Atkinson, C. A., & Lam, W. A. (2024). What's in a name? Experimental evidence of gender bias in recommendation letters generated by ChatGPT. *Journal of Medical Internet Research*, *26*, e51837.

Kollmayer, M., Schober, B., & Spiel, C. (2018). Gender stereotypes in education: Development, consequences, and interventions. *European journal of developmental psychology, 15*(4), 361-377.

Li, X., & Zaki, R. (2024). Harnessing the power of digital resources in mathematics education: The potential of augmented reality and artificial intelligence. In S. Papadakis (Ed.), *IoT, AI and ICT for educational applications: Technologies to Enable Education for All* (pp.191–223). Springer.

Lithner, J. (2000). Mathematical reasoning in task solving. *Educational Studies in Mathematics*, *41*(2), 165–190. http://www.jstor.org/stable/3483188.

Paul, R. & Elder, L. (2019). *The Thinker's Guide to Socratic Questioning.* Rowman & Littlefield.

Piaget, J. (1971). *Psychology and Epistemology: Towards a Theory of Knowledge.* Grossman.

Richardson, M., Golding, J., Isaacs, T., Barnes, I., Wilkinson, D., Swensson, C., & Maris, R. (2025). *Trends in International Mathematics and Science Study (TIMSS) 2023: National report for England.* Department for Education.

Taber, K. S. (2018). Scaffolding learning: Principles for effective teaching and the design of classroom resources. In M. Abend (Ed.), *Effective teaching and learning: Perspectives, strategies and implementation* (pp. 1–43). Nova Science Publishers.

Taber, K. S. (2017). The Role of New Educational Technology in Teaching and Learning: A Constructivist Perspective on Digital Learning. In A. Marcus-Quinn & T. Hourigan (Eds.), *Handbook on Digital Learning for K-12 Schools* (pp. 397-412). Springer International Publishing.

Taber, K. S. (2014). *Student Thinking and Learning in Science: Perspectives on the Nature and Development of Learners' Ideas.* Routledge.

Taber, K. S. (2009). *Progressing Science Education: Constructing the scientific research programme into the contingent nature of learning science*. Springer.

Taber, K. S., & Li, X. (in press). Educational technology in the service of science teaching and learning. In B. Akpan & B. Svendsen (Eds.), *The Convergence of Learning Sciences and Science Education.* Springer Nature.

Taber, K. S., & Li, X. (2021). The vicarious and the virtual: A Vygotskian perspective on digital learning resources as tools for scaffolding conceptual development. In A. M. Columbus (Ed.), *Advances in psychology research* (Vol. 143). Nova Science Publishers.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. Jolm-Steiner, S. Scribner & E. Souberman, Eds.). Harvard University Press.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89–100. http://doi.org/10.1111/j.1469–7610.1976.tb00381.x.