What Kind of Mathematics Teacher is ChatGPT? Identifying the Pedagogical Practices Preferenced by Generative AI Tools When Preparing Lesson Plans

Scott Cameron
University of Melbourne
scameron@unimelb.edu.au

Carmel Mesiti
University of Melbourne
cmesiti@unimelb.edu.au

In their daily work teachers are responsible for several complex tasks; might AI be harnessed to support teachers in the challenging work of planning lessons? In this paper we investigate the use of an AI tool, namely ChatGPT, to generate a lesson plan that may be of use to teachers in their planning. A carefully worded prompt, informed by research, was used to generate four lesson plans for the teaching of division of fractions to students in years 7 or 8. We analysed the plans' structures and encoded practices. AI-generated lesson plans appeared suitable for identifying *what* should be taught but lacked detail of practices that support teachers to teach meaningfully.

Generative Artificial Intelligence (AI) tools (e.g., OpenAI ChatGPT, Google Gemini) have the potential to revolutionise teaching and learning through their capacity to generate a range of text types in a matter of moments (Sabzalieva & Valentini, 2023). When considering educational contexts, AI can be assigned a range of roles. AI can act as a 'guide on the side' when used to support teachers in the generation of classroom materials and advise on the learning sequence of specific concepts. AI can also act as a 'co-designer' of teaching materials when prompted to provide input into the design of curriculum materials (Sabzalieva & Valentini, 2023). Given the significant workload associated with the development of teaching materials (Hunter & Sonnemann, 2022), and the desire from teachers to have more time for lesson planning rather than using stock lesson plans (Stacey et al., 2023), the use of AI to generate teaching materials, including lesson plans, may be beneficial to teachers. We posit that teachers might be able to develop a lesson plan efficiently through using AI to develop a draft that is then refined for use in their classroom.

Although AI can generate a range of materials to support teaching and learning, the classroom teacher remains the expert in selecting instructional materials for use in their class. Often, teachers' choices are based on their pedagogical alignment with the content or focus of the instructional materials (Remillard, 2005). However, AI tools may be likened to a 'black box' where users cannot see or understand how outputs were generated, or the choices that resulted in a given output (Bearman & Ajjawi, 2023), thus potentially making it difficult for teachers to identify the pedagogies that informed the AI output. This paper reports a preliminary attempt to identify the preferred pedagogical practices of a generative AI tool.

Literature Review

Working with AI

AI tools use large language models to learn from a range of training material to complete tasks as requested by an end user. In the case of ChatGPT, the model has been trained on a range of material that is either publicly available on the internet, licenced from third parties, or provided by users and human trainers (Open AI, n.d.). Given the range of publicly available online teaching resources, we expect that this training material has captured an extensive array of lesson plans, curricula, and teaching materials. By analysing these data and identifying word associations, AI tools can generate text in response to a request by a consumer.

(2024). In J. Višňovská, E. Ross, & S. Getenet (Eds.), *Surfing the waves of mathematics education. Proceedings of the 46th annual conference of the Mathematics Education Research Group of Australasia* (pp. 135–142). Gold Coast: MERGA.

Generative AI tools produce text in response to a stimulus. The output contains "a judgement about an optimal course of action" (Bearman & Ajjawi, 2023, p. 1160) informed by the content of the training material. Given the complexity of the algorithms that result in such outputs, Bearman and Ajjawi argue that the decisions that underpin this course of action cannot be observed by the end user. Consequently, it is left to the consumer to develop an approach for working with AI tools. This view is shared by the Commonwealth of Australia (2023) who recognise the role of teacher expertise in using generative AI tools to support and enhance teaching and learning; teachers (and not AI) are recognised as the subject matter experts within the classroom. When working with AI, teachers will need to familiarise themselves with its potential affordances and constraints before they are able to capitalise its use (Su & Yang, 2023). This familiarisation involves new knowledge, such as the development of effective prompts for different tasks, and critical skills for interpreting and refining AI outputs (Commonwealth of Australia, 2023). The establishment of an evidence base, critical in supporting the work of teachers in using AI, is currently lacking across a range of educational contexts (Su & Yang, 2023). The preliminary work reported in this paper aims to assist teachers in understanding and interpreting AI-generated lesson plans.

Teachers as Pedagogical Experts

Alexander (2008) defines pedagogy as the *act* of teaching and connects this practice with discourses of educational theories, values, and evidence, "It is what one needs to know, and the skills one needs to command, in order to make and justify the many different kinds of decisions of which teaching is constituted" (p. 47). Shulman (1987) agrees that expert teaching is characterised by careful management of students and of *ideas* within classroom discourse. Together, these descriptions of teaching practice highlight the complexity of classroom practices where educational objectives are mitigated by specific skills and knowledge that are necessary requirements for the act of teaching. Alexander's metaphorical description of pedagogy as a "deep pool" (2015, p. 253) recognises the inherent challenge of capturing and defining these classroom practices that enable and support both teaching and learning processes.

Teachers develop sophisticated practices that represent the accumulated 'wisdom' of their professional experiences in the classroom (Shulman, 1987). We understand that this pedagogical expertise is formed, refined and powerfully influenced by one's experiences as learners (Schweisfurth, 2015), as teachers (Shulman 1987), and the professional practical experience in training developed to strengthen novices understanding of the nexus between theory and the practice of teaching (Darling-Hammond, 2017).

In contrast, the expected practices evident in AI generated lesson plans will be drawn from existing text-based resources and will be detached from any experiential learning. Consequently, there may be a disconnect between the pedagogical practices of AI and the pedagogical practices of an experienced classroom teacher.

Research Design

The following research question drove the preliminary research presented in this paper:

- What pedagogical practices are preferenced by ChatGPT in the development of lesson plans relating to the division of two fractions?

Rationale for Choice of AI Tool and Lesson Focus

ChatGPT was selected as the AI tool for this study as it is freely available and accessible to teachers. ChatGPT continues to be the most frequently visited AI tool on the internet (Similarweb, n.d.), thus suggesting that ChatGPT would be the most likely AI tool of choice if teachers were to use AI tools to support their lesson planning.

Division of fractions was chosen as the context for this preliminary research as this area of the curriculum is considered difficult for students to learn, and challenging for teachers to teach (e.g., Siemon et al., 2015). Consequently, division of fractions presents as a suitable topic for this analysis as teachers are more likely to draw upon a range of resources to develop their teaching ideas and pedagogical content knowledge; AI presents as one such resource.

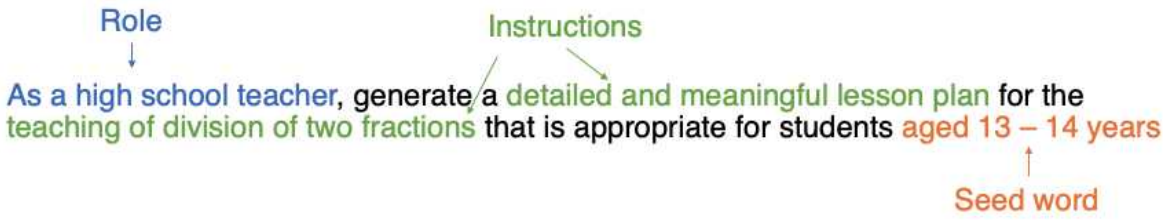
Data Generation

We used ChatGPT to prepare four lesson plans on the division of fractions. ChatGPT explains that its responses "can vary based on a variety of factors, including the context of our conversation, the details you provide in your questions, and any updates or changes to my training data" (OpenAI, 2024a). To account for this potential variation, the same prompt was used, and each lesson plan was requested in a different conversation. Lesson plans were generated on different days to allow for any differences in updates or training data to become apparent as it was expected that these changes may shift the preferred practices of ChatGPT.

Preliminary research by Spasić & Janković (2023) has explored how the structure of the input prompt shapes the quality of the generated lesson plan. Informed by their findings, we developed a prompt (see Figure 1) which specifies a **role** for ChatGPT and includes **instructions** and **seed words**. The inclusion of a role, instructions and seed words results in the generation of a lesson plan that is more detailed than if any of the three elements of the prompts are absent.

Figure 1

Prompt Used to Generate Lesson Plans



Coding of Lesson Plans

Four lesson plans were generated using the prompt in Figure 1. We used findings from The International Classroom Lexicon Project (Mesiti, Artigue, et al., 2021) that identified a set of pedagogical terms, the Australian Lexicon, which teachers use to describe the practices of mathematics classrooms (Mesiti, Hollingsworth, et al., 2021). When offered the 61 terms from the Australian Lexicon, 52 Victorian mathematics teachers were asked to reduce the terms to 'ten essential terms' (Mesiti et al., 2019). The responses were sorted according to frequency and a set of 15 terms were identified: *Assessment*, *Demonstrating*, *Differentiating*, *Engaging*, *Feedback*, *Formative Assessment*, *Group Discussion*, *Group Work*, *Modelling*, *Practising*, *Questioning*, *Reasoning*, *Reflecting*, *Scaffolding* and *Worked Example*. Each of these pedagogical terms have been operationalised with a short description, examples, and non-examples elsewhere (Mesiti et al., 2021b). These 15 terms were adopted as codes for the AI-generated lesson plans. The lesson plans that were generated in response to the prompt listed general 'instructions' under organisational headings. For example, the 'Warm Up' for Lesson Plan 1 stated:

- Review the concept of multiplying fractions briefly;
- Ask students to solve a multiplication problem involving fractions, such as $\frac{2}{3} \times \frac{3}{4}$;
- Discuss the steps involved in multiplying fractions (OpenAI, 2024b).

For two of the lesson plans, each instruction was coded by both researchers as they negotiated their understanding of the pedagogical terms. The other two lessons were coded independently. Coding was determined by the alignment of text from the lesson plan with the operationalised definition of the terms. For example, "Ask students to solve a multiplication problem involving fractions" was coded as Questioning and Practising.

Results and Discussion

The following sections report the results of two analyses related to the structure of, and pedagogical practices evident in, the four AI-generated lesson plans.

Lesson Plan Structure

The structure of each of the four lesson plans, determined from the section headings from the generated responses, is summarised in Table 1.

Table 1

Lesson Structure Headings in AI-Generated Lesson Plans

Heading	Lesson plan 1	Lesson plan 2	Lesson plan 3	Lesson plan 4
Pre-lesson plan headings		Title	Topic	
		Grade level	Grade level	
		Subject	Subject	
		Duration	Duration	
	Objective	Objective	Objective	Objective
	Materials needed	Materials	Materials needed	Materials needed
Lesson plan headings	Warm up	Warm up		
	Introduction to division of fractions	Introduction	Lesson introduction	Introduction
				Reciprocal fractions
		Direct instruction		Example problems
	Guided practice	Guided practice	Guided practice	Guided practice
	Independent practice	Independent practice	Independent practice	Independent practice
	Real-world application		Real-world application	
	Closure	Closure	Closure	Closure
Post-lesson plan headings	Extension activity	Extension (optional)	Extension	Extension activity (optional)
	Assessment	Assessment	Assessment	Assessment
	Differentiation	Differentiation	Differentiation	Differentiation
			Integration	
	Homework	Homework		Homework
	Reflection			Reflection

While the AI lesson plans appear to adhere to a structural formula, particularly in the pre- and post-lesson sections, there is some variation in the sequence and content of the main body of the lesson plan. The provision of *objectives* aligns with the advice to teachers about the importance of establishing and providing learning goals that define for students the purpose of the lesson and desired achievements (e.g., State of Victoria, 2017). Where a *warm up* was provided, these focussed on a review of multiplication of fractions, thus incorporating the practice of connecting to prior knowledge which has been shown to support learning (Lovitt & Clarke, 2011). Common to all lesson plans were *introductions* focussed on reviewing the

concept of fractions and/or division. For example, particular emphasis was given to division of whole numbers in Lesson Plan 4 (LP4), and conceptual understanding of division in LP3. In contrast, LP1 and LP2 focussed on explaining how fraction division is similar to multiplication, albeit with a "slight twist" (LP2).

All four lesson plans outline a step-by-step approach for the division of two fractions, namely, the common rule 'invert and multiply' (Davis & Pearn, 2009). This approach was included in the *introduction* (for LP1 and LP3), in a section named *direct instruction* (for LP2), and in a section named *reciprocal fractions* (for LP4). Similarly, the inclusion of *example problems* in LP4 is encapsulated within the *guided practice* sections of the other lesson plans. These two examples suggest a default in the pedagogical approaches generated by ChatGPT despite the different language expressions and structures of the lesson plans. The inclusion of *example problems* and *guided practice* is consistent with the identification of worked examples as an important teaching strategy (e.g., State of Victoria, 2017) and reflects research findings that highlights the common use of worked examples in the mathematics classroom (Große, 2015). Notably, only LP1 and LP4 specified the examples that should be used ($\frac{2}{3} \div \frac{1}{4}$ in LP1; $\frac{1}{2} \div \frac{1}{4}$ in LP4). The absence of several worked examples in these lesson plans, possibly a ChatGPT default, indicates a significant aspect of the lesson plan which would require teachers to draw upon their expertise and understanding of the affordances of worked examples (e.g., Chick, 2007) before delivering the lesson. Indeed, the example offered to illustrate division of fractions in LP1, is not necessarily an example that would be used by an expert teacher in the very first instance (see also Davis & Pearn, 2009).

The inclusion of post-lesson materials was unprompted, and the consistency of structure again suggests a standard approach for the generation of lesson plans. In all lesson plans, the only evidence of *extension* and *differentiation* is in these post-lesson sections.

Practices Evident in the Lesson Plans

Table 2 and Figure 2 summarise the results of the coding of the four lesson plans against the *essential* terms of the Australian Lexicon, offering insight into the pedagogical practices preferenced in the development of these lesson plans. The most frequent code, *Group Discussion* ($n = 36$), reflects a significant proportion of instructions in the lesson plans that require either a whole classroom discussion or a teacher explanation. The instructions in the lesson plans are brief and do not identify an agent, so we were required to anticipate the most likely classroom activity and setting. Accordingly, it was assumed that such discussions and explanations would involve *Reasoning*, contributing to its high frequency ($n = 25$). Notwithstanding these, included in the lesson plans were specific instances where *Reasoning* was identified as the key code, "Encourage students to explain their reasoning as they solve each problem" (LP2). The frequency of codes *Demonstrating* ($n = 13$), *Modelling* ($n = 11$) and *Questioning* ($n = 15$), align with a pedagogical preference for the use of 'explicit teaching', a High Impact Teaching Strategy (State of Victoria, 2017), for classroom instruction of division of fractions.

Practising, the third most coded essential term ($n = 23$), identifies student activity. Instructions coded against this essential term involved instructing students to complete problems, as well as those classroom activities where both the teacher and the students complete problems together. The appearance of this practice is consistent with the notion of 'collaborative learning'; another High Impact Teaching Strategy (State of Victoria, 2017).

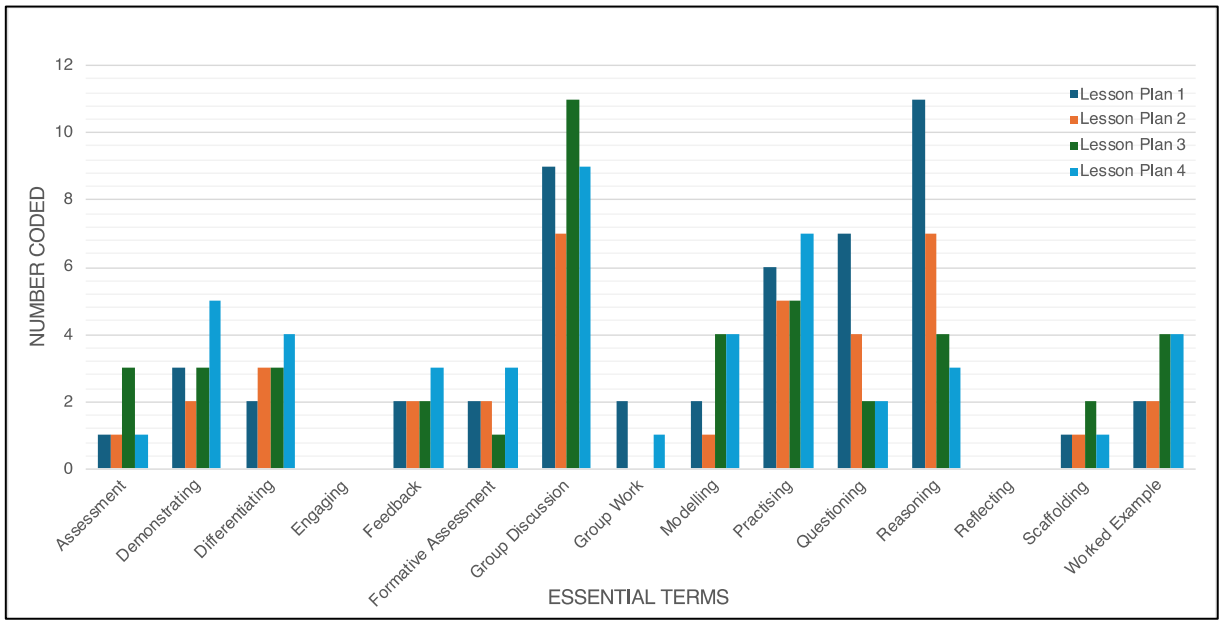
Table 2

Number of Essential Terms Coded in AI-Generated Lesson Plans

Essential terms	Lesson plan 1 (391 words)	Lesson plan 2 (298 words)	Lesson plan 3 (367 words)	Lesson plan 4 (406 words)	Total
Assessment	1	1	3	1	6
Demonstrating	3	2	3	5	13
Differentiating	2	3	3	4	12
Engaging	0	0	0	0	0
Feedback	2	2	2	3	9
Formative Assessment	2	2	1	3	8
Group discussion	9	7	11	9	36
Group work	2	0	0	1	3
Modelling	2	1	4	4	11
Practising	6	5	5	7	23
Questioning	7	4	2	2	15
Reasoning	11	7	4	3	25
Reflecting	0	0	0	0	0
Scaffolding	1	1	2	1	5
Worked example	2	2	4	4	12
Total	50	37	44	47	178

Figure 2

Number of Essential Terms Coded in AI-Generated Lesson Plans



Notably all but three *essential* terms were present in all lesson plans (*Engaging*, *Reflecting* and *Group Work*). The practice of *Engaging*, defined as "A student is actively involved with an educational experience, whereby he/she acts to maintain or extend their contact with the stimulus (typically, in order to increase their knowledge of it)" (Mesiti et al., 2021b, p. 44), is likely perceptible from teacher/student and student/student interactions rather than from lesson plans documenting the scope and structure of a lesson. The practice of *Reflecting*, defined as

"An activity in which students consider the effectiveness or progress of their learning (i.e., their developing knowledge, skills, and understandings)" (p. 49), was also absent, despite all lesson plans including a *closure*. The instructions related to the sections named *closure* generally involved reviewing and summarising what was covered in the lesson and assigning of homework. *Group Work,* whereby "Students work together to complete a given activity" (p. 45) was not evident in two of the four lesson plans. This absence may be due to the brevity of the lesson plans (ranging from 298 to 406 words) and although the instructions included in the lesson plans specified *what* should be taught, the *how* of teaching was more difficult to discern.

Conclusion

This paper sought to identify the pedagogical practices favoured by the AI tool, ChatGPT, when tasked with generating a lesson plan. The use of the fifteen *essential terms* (Mesiti et al., 2019), a subset of the Australian Lexicon (Mesiti, Hollingsworth, et al., 2021), proved useful in providing a framework through which the AI-generated lesson plans could be coded, and aided in the identification of pedagogical practices present and preferenced in the four lesson plans. Further analysis using the 61 terms of the entire lexicon would enable a more detailed description of the pedagogical practices preferenced in ChatGPT outputs and, we expect, further highlight the important role of teacher expertise in refining and implementing AI-generated lesson plans. It is worth noting the impact of prompts on outputs. Further research could explore the impact of different prompts on the development of lesson plans and provide guidance to teachers so they may develop effective prompts that support the generation of a lesson plan with greater detail with respect to explanations, examples, and suitable problems.

Despite slight differences in structure, the four lesson plans favoured the following pedagogical approach: key procedures or concepts are introduced, key steps and skills are demonstrated and illustrated with worked examples, and additional problems are set for students to complete. The key concept of 'division of fractions' was illustrated with the 'invert and multiply approach' separated from any meaningful representation. While this approach may be suitable in some contexts, it reflects a narrow view of the potential for mathematics teaching by appearing to favour traditional 'telling', stating of information or demonstrating of procedures (Smith, 1996), and practices for developing deep thinking, reflection and justification by students, remain absent. While several practices align with evidence-based approaches for teaching, these practices were general and did not incorporate evidence-based approaches that can support the learning of division of fractions (e.g., the use of a bar model; Yeap, 2011). The absence of such mathematical approaches, along with the absence of specific examples, explanations, and problems highlight the considerable expertise needed by teachers to refine and implement AI-generated lesson plans. In the case of lesson plans generated by ChatGPT, advice is provided for teachers on *what* to teach, but not *how* to teach it.

References

Alexander, R. (2008). *Essays on pedagogy.* Routledge. https://doi.org/10.4324/9780203609309

Alexander, R. J. (2015). Teaching and learning for all? The quality imperative revisited. *International Journal of Educational Development, 40,* 250–258. https://doi.org/10.1016/j.ijedudev.2014.11.012

Bearman, M., & Ajjawi, R. (2023). Learning to work with the black box: Pedagogy for a world with artificial intelligence. *British Journal of Educational Technology*, *54*(5), 1160–1173. https://doi.org/10.1111/bjet.13337

Chick, H. (2007). Teaching and learning by example. In J. Watson, & K. Beswick (Eds.), *Mathematics: Essential research, essential practice. Proceedings of the 30th annual conference of the Mathematics Education Research Group of Australasia* (pp. 3–21). Hobart: MERGA.

Commonwealth of Australia (2023). *Australian framework for generative artificial intelligence (ai) in schools.* Commonwealth of Australia. https://www.education.gov.au/download/17416/australian-framework-generative-artificial-intelligence-ai-schools/35400/australian-framework-generative-ai-schools/pdf

Darling-Hammond, L. (2017). Teacher education around the world: What can we learn from international practice? *European Journal of Teacher Education*, *40*(3), 291–309. https://doi.org/10.1080/02619768.2017.1315399

Davis, G. E., & Pearn, C. A. (2009). *Division of fractions.* Republic of Mathematics.

Große, C. S. (2015). Fostering modeling competencies: Benefits of worked examples, problems to be solved, and fading procedures**.** *European Journal of Science and Mathematics Education*, *3*(4), 364–375. https://doi.org/10.30935/scimath/9444

Hunter, J., & Sonnemann, J. (2022). *Making time for great teaching: How better government policy can help.* Grattan Institute. https://grattan.edu.au/report/making-time-for-great-teaching-how-better-government-policy-can-help/

Lovitt, C., & Clarke, D. (2011). The features of a rich and balanced mathematics lesson: Teacher as designer. *Educational Designer, 1*(4), Article 15, 1–25.

Mesiti, C., Artigue, M., Hollingsworth, H., Cao, Y., & Clarke, D. J. (Eds.). (2021a). *Teachers talking about their classrooms: Learning from the professional lexicons of mathematics teachers around the world.* Routledge. https://doi.org/10.4324/9780429355622

Mesiti, C., Clarke, D. J., & van Driel, J. (2019). Describing and prescribing classroom practice: Do we have a common language? In G. Hine, S. Blackley, & A. Cooke (Eds.), *Mathematics. Proceedings of the 42nd annual conference of the Mathematics Education Research Group of Australasia* (pp. 492–499). MERGA.

Mesiti, C., Hollingsworth, H., Clarke, D. J., Reed, A.S., & Roan, K. (2021b). Australian Lexicon. In C. Mesiti, M. Artigue, H. Hollingsworth, Y. Cao, & D. J. Clarke (Eds*.), Teachers talking about their classrooms: Learning from the professional lexicons of mathematics teachers around the world* (pp. 39–54). Routledge. https://doi.org/10.4324/9780429355622-3

OpenAI (2024a). *ChatGPT* (Feb 16 version) [Large language model]. https://chat.openai.com/chat

OpenAI (2024b). *ChatGPT* (Feb 18 version) [Large language model]. https://chat.openai.com/chat

OpenAI (n.d.). *How ChatGPT and our language models are developed.* https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed

Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research*, *75*(2), 211–246. https://doi.org/10.3102/00346543075002211

Sabzalieva, E., & Valentini, A. (2023). *ChatGPT and artificial intelligence in higher education: Quick start guide*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000385146

Schweisfurth, M. (2015). Learner-centred pedagogy: Towards a post-2015 agenda for teaching and learning. *International Journal of Educational Development*, *40*, 259–266. https://doi.org/10.1016/j.ijedudev.2014.10.011

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1–23. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Similarweb (n.d.). *Top websites ranking*. https://www.similarweb.com/top-websites/

Smith, J. P. (1996). Efficacy and teaching mathematics by telling: A challenge for reform. *Journal for Research in Mathematics Education*, *27*, 387–402. https://doi.org/10.5951/jresematheduc.27.4.0387

Spasić, A. L., & Janković, D. S. (2023). Using ChatGPT standard prompt engineering techniques in lesson preparation: Role, instructions and seed-word prompts. In *58th international scientific conference on information, communication and energy systems and technologies*. IEEE. https://doi.org/10.1109/ICEST58410.2023.10187269

Stacey, M., Gavin, M., Fitzgerald, S., McGrath-Champ, S., & Wilson, R. (2023). Reducing teachers' workload or deskilling 'core' work? Analysis of a policy response to teacher workload demands. *Discourse: Studies in the Cultural Politics of Education*, *45*(2), 189–199. https://doi.org/10.1080/01596306.2023.2271856

State of Victoria (2017). *High impact teaching strategies; Excellence in teaching and learning*. Department of Education and Training, Victoria.

Su, J., & Yang, W. (2023). Unlocking the power of ChatGPT: A framework for applying generative AI in education. *ECNU Review of Education*, *6*(3). https://doi.org/10.1177/20965311231168423

Yeap, B.H. (2011). *Bar modeling–A problem-solving tool: From research to practice (an effective Singapore math strategy)*. Marshall Cavendish Education.